Numerical Methods for Solving
Partial Differential Equations
Numerical Methods for Solving Partial Differential Equations

A Comprehensive Introduction for Scientists and Engineers

George F. Pinder

WILEY
Robert N. Farvolden and John D. Bredehoeft
my mentors
## Contents

**Preface** \hspace{1cm} xi

1 **Interpolation** \hspace{1cm} 1
   1.1 Purpose \hspace{1cm} 1
   1.2 Definitions \hspace{1cm} 1
   1.3 Example \hspace{1cm} 2
   1.4 Weirstraus Approximation Theorem \hspace{1cm} 3
   1.5 Lagrange Interpolation \hspace{1cm} 4
      1.5.1 Example \hspace{1cm} 6
   1.6 Compare \( P_2(\theta) \) and \( \hat{f}(\theta) \) \hspace{1cm} 8
   1.7 Error of Approximation \hspace{1cm} 9
   1.8 Multiple Elements \hspace{1cm} 14
      1.8.1 Example \hspace{1cm} 17
   1.9 Hermite Polynomials \hspace{1cm} 20
   1.10 Error in Approximation by Hermites \hspace{1cm} 23
   1.11 Chapter Summary \hspace{1cm} 24
   1.12 Problems \hspace{1cm} 24

2 **Numerical Differentiation** \hspace{1cm} 33
   2.1 General Theory \hspace{1cm} 33
   2.2 Two-Point Difference Formulae \hspace{1cm} 34
      2.2.1 Forward Difference Formula \hspace{1cm} 35
      2.2.2 Backward Difference Formula \hspace{1cm} 36
      2.2.3 Example \hspace{1cm} 36
      2.2.4 Error of the Approximation \hspace{1cm} 36
   2.3 Two-Point Formulae from Taylor Series \hspace{1cm} 37
   2.4 Three-point Difference Formulae \hspace{1cm} 40
      2.4.1 First-Order Derivative Difference Formulae \hspace{1cm} 41
      2.4.2 Second-Order Derivatives \hspace{1cm} 43
   2.5 Chapter Summary \hspace{1cm} 46
   2.6 Problems \hspace{1cm} 46

3 **Numerical Integration** \hspace{1cm} 55
   3.1 Newton-Cotes Quadrature Formula \hspace{1cm} 55
      3.1.1 Lagrange Interpolation \hspace{1cm} 55
      3.1.2 Trapezoidal Rule \hspace{1cm} 56
      3.1.3 Simpson’s Rule \hspace{1cm} 57
      3.1.4 General Form \hspace{1cm} 58
      3.1.5 Example using Simpson’s Rule \hspace{1cm} 59
11 Boundary-Value Problems in Three Space
   11.1 Finite Difference Approximations ................................ 279
   11.2 Finite Element Approximations .................................... 280
   11.3 Chapter Summary ..................................................... 285

12 Nomenclature ............................................................. 289

Index ................................................................................. 293
Preface

While there are many good books on numerical methods suitable for students of mathematics and many others that are accessible to scientists and engineers, but dedicated to a specific discipline; there is a need for a book that is accessible to students of science and engineering that is not discipline specific, yet rigorous and comprehensive in scope. This book is an effort to fill this need.

Herein I provide the logical underpinnings of all of the commonly encountered numerical methods, namely finite difference, finite element, collocation, and finite volume methods, at a level of sophistication consistent with the needs and interests of science and engineering students. Two mathematical concepts, namely polynomial approximation theory and the method of weighted residuals, form the intellectual framework for the introduction and explanation of all of these methods.

The approach is to first introduce polynomial approximation theory in one space dimension followed by the concept of the methods of weighted residuals. Employing only polynomial approximation theory the finite difference method is easily developed and presented. With the addition of the method of weighted residuals, finite element, collocation and finite volume methods are readily accessible. These concepts are introduced first in one space dimension, then the time dimension, then two space dimensions, and finally two space dimensions and time.

The equations considered are first order, second order, and second order in space and first order in time. By design, the book does not focus on any specific area of science or engineering. It is designed to teach numerical methods as a concept rather than as applied to a specific discipline. The intent is to provide the student with the ability to understand numerical methods as encountered in technical readings specific to his/her discipline and to be able to apply them in practice.

The book assumes a knowledge of matrix algebra and differential equations. A programming language is also needed if the reader is interested in applying numerical methods to example problems. No prior knowledge of numerical methods is assumed. While a few theorems are used, no proofs are presented.

This book stems from a course I teach in Numerical Methods for Engineers. The course is taught as a precept and typically populated by an approximately equal number of senior undergraduates and graduate students from different engineering disciplines. A project of practical significance is assigned that requires the creation of a computer program capable of solving a second-order two-space dimensional equation using finite elements.

I am indebted to Xin Kou, my doctoral student in mathematics, for carefully reviewing the manuscript for his book, identifying notational inconsistencies and making important suggestions as to how to improve the presentation.
Chapter 1

Interpolation

In this chapter, we will introduce interpolation theory, the first of two key topics that will form the foundation of everything that follows in this book. We will find that this concept leads quite naturally to finite difference methods and, when combined with the second key topic, the method of weighted residuals, provides the necessary mathematical concepts needed for all other numerical methods we present. So, let us get started.

1.1 Purpose

Interpolation is a method of constructing new data points between known values or for creating a function that fits exactly a known set of discrete data points defined within a specific range. Interpolation has many applications in science and engineering. In this book, it will be used to form the basis for numerical differentiation, numerical quadrature, numerical integration, and as a key part of several numerical methods used to solve differential and partial differential equations.

1.2 Definitions

We begin by introducing some interpolation notation. Consider a region

\[(a =) \quad x_0 < x_1 < x_2 \ldots < x_n (= b)\]  \hspace{1cm} (1.1)

as illustrated in Fig. 1.1

![Discretized line spanning a to b.](image)

Next assume there exists a function \(f(x)\) that is a known function of \(x\). We will use this function momentarily.

Now also consider a function \(P_n(x)\) that has the following properties:
Table 1.1: Comparison of exact and calculated approximations for the sin curve.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$f(\theta)$</th>
<th>$P_2(\theta)$</th>
<th>$E(\theta) \equiv f(\theta) - P_2(\theta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\pi/4$</td>
<td>0.707</td>
<td>0.753</td>
<td>-0.046</td>
</tr>
<tr>
<td>$\pi/2$</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>$3\pi/4$</td>
<td>0.707</td>
<td>0.751</td>
<td>-0.044</td>
</tr>
<tr>
<td>$\pi$</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$5\pi/4$</td>
<td>-0.707</td>
<td>-1.18</td>
<td>0.37</td>
</tr>
<tr>
<td>$6\pi/4$</td>
<td>-1</td>
<td>-2.89</td>
<td>1.89</td>
</tr>
</tbody>
</table>

1. $P_n(x)$ is a polynomial of degree $n$, that is,

$$P_n(x) = \sum_{i=0}^{n} a_i x^i$$  \hspace{1cm} (1.2)

where the coefficients $a_i$ are known constants and $x^i$ indicates the variable $x$ to the $i$th power.

2. $P_n(x_i) = f(x_i)$ where $x_i$ are particular values of $x$ as seen in Fig. 1.1. In other words at the locations $x_i$ the values of $f(x)$ and $P_n(x)$ are identical.

According to our definition of interpolation, this $P_n(x)$ is an interpolating polynomial of degree $n$. Note that by convention, $i$ has a lower bound value of 0 rather than 1. To make the above clear, consider the following example.

### 1.3 Example

Consider the sine function shown in Fig. 1.2. and the information presented in Table 1.1.

![Figure 1.2: Sin ($\theta$) curve with measured points indicated by small black dots (dashed curve) and interpolated values indicated by large black dots (solid curve).]
Let the second-degree polynomial \( P_2(\theta) \) be given by

\[
P_2(\theta) = a_0 + a_1\theta + a_2\theta^2.
\]  

(1.3)

Now create the special version of Eq. (1.3) that satisfies the three known values of \( P_2(\theta) \), namely those at \( \theta = 0, \pi/2, \) and \( \pi \). Then, substituting values of \( P_2(\theta) \) from Table 1.1 we obtain

\[
P_2(0) = a_0 + a_10 + a_20^2 = 0
\]

\[
P_2(\pi/2) = a_0 + a_1(\pi/2) + a_2(\pi/2)^2 = 1
\]

\[
P_2(\pi) = a_0 + a_1(\pi) + a_2(\pi)^2 = 0
\]

(1.4)

from which we can generate the set of equations

\[
\begin{bmatrix}
1 & 0 & 0 \\
1 & \pi/2 & (\pi/2)^2 \\
1 & \pi & (\pi)^2
\end{bmatrix}
\begin{bmatrix}
a_0 \\
a_1 \\
a_2
\end{bmatrix}
= 
\begin{bmatrix}
0 \\
1 \\
0
\end{bmatrix}.
\]

Solving for the coefficients \( a_0, a_1, \) and \( a_2 \) we obtain

\[
P_2(\theta) = 0 + 1.27(\theta) - 0.40(\theta^2).
\]  

(1.5)

From this expression, we can obtain values of \( P(\theta) \) for any \( \theta \). In Table 1.1, we calculate the values of \( P(\theta) \) for various values of \( \theta \). Notice the difference in the error of the interpolation at \( \theta = \pi/4 \) and \( \theta = (5\pi)/4 \). Why did this happen? It is due to the fact that the value \( \theta = 5\pi/4 \) lies outside of the range of \( \theta \) used to define \( P(\theta) \).

The question now arises as to whether polynomials can be used to represent functions other than the sine. To answer this question we turn to the Weirstraus approximation theorem.

### 1.4 Weirstraus Approximation Theorem

The Weirstraus approximation theorem basically tells us that it is possible to calculate a polynomial approximation of any desired accuracy, provided you employ a suitably large number of terms in the polynomial. It states:

**Theorem 1** If \( f(x) \) is continuous on a finite interval \([a, b]\), then, given any \( \varepsilon > 0 \), there exists an \( n \) and a polynomial \( P_n(x) \) of degree \( n \) such that \( |f(x) - P_n(x)| < \varepsilon \) for all \( x \) in \([a, b]\).

Although this theorem indicates that a polynomial can be found to represent any function, it does not mean that the coefficients of all polynomials can be calculated. In some cases, especially for large \( n \), the coefficient matrix for the polynomial coefficients can be ill-conditioned (almost singular) and the coefficient values unobtainable. Fortunately, in our work, we will be using polynomials that do not exhibit this pathology.
1.5 Lagrange Interpolation

Let us now focus on one special kind of interpolation that we will use extensively in subsequent material. Consider the approximation of a function \( f(x) \) denoted as \( \hat{f}(x) \) written as follows:

\[
    f(x) = \hat{f}(x) + E(x)
\]

where we will call \( E(x) \) the error of the approximation; in other words \( E(x) \) is a measure of how well \( \hat{f}(x) \) represents \( f(x) \).

We now define the form of \( \hat{f}(x) \) in a very special way, that is

\[
    \hat{f}(x) = \sum_{j=0}^{n} \ell^n_j(x) f(x_j)
\]

where \( \ell^n_j(x) \) is an as yet undefined polynomial of degree \( n \).

To determine the functional form of the polynomial \( \ell^n_j(x) \) we will require that \( E(x_i) = 0 \), the error in the approximation, vanishes at selected locations along \( x \), namely at \( x_i, i = 0, \ldots, n \). We will call these locations nodes and they are indicated in Fig. 1.1 by the black dots. Recall that we required \( f(x) \) to equal \( f(x_j) \) in our general definition of a polynomial in Section 1.2. We can write this requirement formally, in terms of the errors \( E(x_i) \), as

\[
    E(x_i) = 0 \quad i = 0, \ldots, n
\]

where we have \( n + 1 \) nodes. Now combine Eq. (1.9) with Eq. (1.8) to give

\[
    f(x_i) = \sum_{j=0}^{n} \ell^n_j(x_i) f(x_j) \quad i = 0, \ldots, n.
\]

Equation (1.10) is Eq. (1.8) written for the specific nodal locations \( x_i \) where, by definition, the error must vanish. Note that the index \( i \) identifies the location, that is \( f(x_i) \) and \( \ell^{(n)}_j(x_i) \) where the polynomial is being evaluated and the index \( j \) indicates the term in the polynomial, that is \( \ell^{(n)}_j(x_i) f(x_j) \). Let us now expand Eq. (1.10) as

\[
    f(x_1) = \ell^n_0(x_1) f(x_0) + \ell^n_1(x_1) f(x_1) + \cdots + \ell^n_n(x_1) f(x_n).
\]

The form of Eq. (1.11) suggests that the polynomials \( \ell^n_j(x_i) \) must have special properties. In order to satisfy the requirement that

\[
    E(x_i) = 0, \quad i = 0, \ldots, n.
\]

Indeed, at the location \( x_1 \), for example, Eq. (1.11) must yield the following:

\[
    \ell^n_1(x_1) = 1 \quad \ell^n_0(x_1) = \ell^n_2(x_1) = \cdots = \ell^n_n(x_1) = 0.
\]

In fact we can generalize this statement to any nodal location \( x_i \), that is
1.5. **LAGRANGE INTERPOLATION**

\[
\ell_j(x_i) = \begin{cases} 
1, & i = j \\
0, & i \neq j 
\end{cases}, \quad i, j = 0, \ldots, n. \tag{1.15}
\]

In other words, our polynomial must be unity at the nodal location for which it is defined, that is where the indices \(i\) and \(j\) are the same, and zero at all other nodes. Writing this in shorthand notation we get that

\[
\ell_j^n(x) = \delta_{ij} \quad j = 0, \ldots, n, \quad i = 0, \ldots, n \tag{1.16}
\]

where \(\delta_{ij}\) is the Kronecker delta. The Kronecker delta is defined such that

\[
\delta_{ij} = \begin{cases} 
1, & i = j \\
0, & i \neq j 
\end{cases} \tag{1.17}
\]

Now let us assume the special case of \(n = 1\); that is we are considering a linear polynomial. Let us focus on the \(j\)th polynomial, that is

\[
\ell_j^1(x) = a_j + b_j x. \tag{1.18}
\]

In light of Eq. (1.17), we can say that the following is true for \(\ell_j^1(x)\) evaluated at node \(x_0\)

\[
\ell_0^1(x_0) = a_0 + b_0 x_0 = 1 \tag{1.19}
\]

\[
\ell_1^1(x_0) = a_0 + b_0 x_1 = 0. \tag{1.20}
\]

Writing this system of equations in matrix form we get

\[
\begin{bmatrix} 1 & x_0 \\ 1 & x_1 \end{bmatrix} \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{1.21}
\]

which we can solve to obtain

\[
\begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \frac{1}{x_1 - x_0} \begin{bmatrix} x_1 & -x_0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{x_1}{x_1 - x_0} \\ \frac{-x_0}{x_1 - x_0} \end{bmatrix}. \tag{1.22}
\]

Substitution of Eq. (1.22) into Eq. (1.18) yields

\[
\ell_0^1(x) = a_0 + b_0 x = \frac{x_1}{x_1 - x_0} - \frac{1}{x_1 - x_0} x = \frac{x_1 - x}{x_1 - x_0}. \tag{1.23}
\]

One can similarly obtain \(\ell_1^1(x) = \frac{x - x_0}{x_1 - x_0}\). The functions \(\ell_0^1(x)\) and \(\ell_1^1(x)\) are **linear Lagrange polynomials** and are presented in Fig. 1.3

Let us check to see if these functions satisfy the requirements stated in Eqs. (1.19) and (1.20):

- when \(x = x_0\): \(\ell_0^1(x_0) = \frac{x_1 - x_0}{x_1 - x_0} = 1 \tag{1.24}\)
- when \(x = x_1\): \(\ell_1^1(x_1) = \frac{x_1 - x_1}{x_1 - x_0} = 0. \tag{1.25}\)

It appears the Lagrange polynomials, as defined above, work for the linear case.

The general form of the \(n\)th degree Lagrange polynomial is

\[
\ell_j^n(x) = \prod_{i=0, \ i \neq j}^n \frac{x - x_i}{x_j - x_i}. \tag{1.26}
\]
where the operator $\Pi^n_{i=0}^{i\neq j}$ says that for a specific value of $j$, each term denoted by the subscript $i$, $i = 0, 1, ..., n$ will be multiplied together except for the special case of $i = j$, which will be skipped because it would result in a value of zero in the denominator of that term. For example, $\Pi^3_{i=1}x_i = (x_1)(x_2)(x_3)$. Thus, for our linear case we obtain

$$\ell^1_0 (x) = \Pi^1_{i=0}^{i\neq j} \frac{x - x_i}{x_0 - x_i} = \frac{x - x_1}{x_0 - x_1} \quad (1.27)$$

which is the same as we obtained in Eq. (1.23) after multiplying the numerator and denominator by $(-1)$.

The strategy that was used above for the linear polynomial can be extended to define the quadratic. In this case, as we will see below, there are three unknown coefficients and therefore one needs three equations. The equations are obtained by imposing the constraints defined in Eq. (1.16) on the quadratic polynomial. Alternatively, since we have stated the general polynomial form in Eq. (1.16), we can write directly, by selecting $n = 2$ the relationship (where $j$ is now equal to 0):

$$\ell^2_0 (x) = \Pi^2_{i=0}^{i\neq j} \frac{x - x_i}{x_0 - x_i} = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} \quad (1.28)$$

The shape of this function is shown in Fig. 1.4. Quadratic Lagrange polynomial identified with node $x_0$. By selecting other values of $j$, that is $j = 1$ or $j = 2$ two additional quadratics will be generated for location $x_1$ and $x_2$.

### 1.5.1 Example

Let us use the linear function to represent the function $\ln(x)$. The linear Lagrange polynomials are presented in Fig. 1.3. There will be two terms, one associated with each node, that is $x = x_0$ and $x = x_1$, as seen in Fig. 1.3. Thus we have that $x \in [1, 2]$

$$\hat{f} (x) = \sum_{j=0}^{1} \ell^1_j (x) f (x_j) \quad x \in [1, 2]. \quad (1.29)$$

From this equation we see that we need $f (x_0)$ and $f (x_1)$. To get this information we need to evaluate $\ln(x)$ at $x = 1.0$ and $x = 2.0$. The following equation shows how this is used:
1.5. LAGRANGE INTERPOLATION

Figure 1.4: Quadratic Lagrange polynomial identified with node \( x_0 \).

<table>
<thead>
<tr>
<th>( x )</th>
<th>( f(x) )</th>
<th>( \hat{f}(x) )</th>
<th>( E(x) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>1.25</td>
<td>0.22</td>
<td>0.173</td>
<td>0.047</td>
</tr>
<tr>
<td>1.50</td>
<td>0.40</td>
<td>0.347</td>
<td>0.053</td>
</tr>
<tr>
<td>1.75</td>
<td>0.56</td>
<td>0.520</td>
<td>0.04</td>
</tr>
<tr>
<td>2.00</td>
<td>0.693</td>
<td>0.693</td>
<td>0.00</td>
</tr>
</tbody>
</table>

Table 1.2: Values of the function \( f(x) = \ln(x) \), the approximation to \( f(x) \), that is \( \hat{f}(x) \) and the error \( E(x) = f(x) - \hat{f}(x) \).

\[
\hat{f}(x) = \ell_0^1(x) f(x_0) + \ell_1^1(x) f(x_1) \\
= \ell_0^1(x) f(1.0) + \ell_1^1(x) f(2) \\
= \ell_0^1(x) (0) + \ell_1^1(x) (0.693) \\
= \frac{x - x_0}{x_1 - x_0} (0.693) \\
= \frac{x - 1}{2 - 1} (0.693) \\
= -0.693 + 0.693x. \quad (1.30)
\]

A comparison of the function \( f(x) \) and the approximation \( \hat{f}(x) \) is presented in Table 1.2. Note that at the node points \( x = 1.0 \) and \( x = 2.0 \) the solution is exact, as required by our definition of the approximating polynomial.

It is helpful to examine the information provided in Fig. 1.5. The interpolant \( \hat{f} \) is given in the top pane. It is a straight line since it is made up of the weighted sum of two straight lines as can be seen from Eq. (1.30). The approximation is the sum of the linear Lagrange polynomial at \( x = 1.0 \) multiplied by the value of \( \ln(1) \) and the linear Lagrange polynomial defined at \( x = 2 \) multiplied by \( \ln(2) \). The weighted sum of linear polynomials always generates a linear polynomial approximation. The lower
pane in this figure shows the comparison between the function \( f(x) = \ln(x) \) and its approximation \( \hat{f}(x) \).

Figure 1.5: Representation of the \( \ln(x) \) function using linear Lagrange polynomials.

### 1.6 Compare \( P_2(\theta) \) and \( \hat{f}(\theta) \)

In this section we want to examine the relationship between the use of a quadratic polynomial, and an approximation based upon quadratic Lagrange polynomials, to interpolate. We start by writing the approximation of the \( \sin(\theta) \) function, \( \hat{f}(x) \), using the quadratic Lagrange polynomials, that is

\[
\hat{f}(\theta) = \sum_{j=0}^{2} \ell_j^2(\theta) f_j(\theta) \tag{1.31}
\]

where, using Eq. (1.28), we obtain three polynomials, one for each node in Fig. 1.4

\[
\ell_0^2(\theta) = \frac{(\theta - \theta_1)(\theta - \theta_2)}{(\theta_0 - \theta_1)(\theta_0 - \theta_2)} \tag{1.32}
\]

\[
\ell_1^2(\theta) = \frac{(\theta - \theta_0)(\theta - \theta_2)}{(\theta_1 - \theta_0)(\theta_1 - \theta_2)} \tag{1.33}
\]

\[
\ell_2^2(\theta) = \frac{(\theta - \theta_0)(\theta - \theta_1)}{(\theta_2 - \theta_0)(\theta_2 - \theta_1)}. \tag{1.34}
\]

Now we multiply each function by the appropriate coefficient value \( f(\theta) \) and get

\[
\hat{f}(\theta) = \ell_0^2(\theta) f(0) + \ell_1^2(\theta) f(1) + \ell_2^2(\theta) f(0). \tag{1.35}
\]

Next we substitute the definitions of \( \ell_1^2(\theta) \)

\[
\ell_1^2(\theta) = \frac{(\theta - 0)(\theta - \pi)}{\pi/2 - 0}(\pi/2 - \pi) = 1.27\theta - 0.406\theta^2 \tag{1.36}
\]

Comparison of this relationship with Eq. (1.5) shows that

\[
P_2(\theta) = \hat{f}(\theta). \tag{1.37}
\]
1.7. ERROR OF APPROXIMATION

In essence, no matter how you manipulate quadratic polynomials, whether or not they are Lagrange polynomials, you will not change the approximation. The reason we use Lagrange polynomials and the \( \tilde{f}(\theta) \) machinery will become more evident later.

As to the approximation, the value of \( \ell^2 \) in Fig. 1.6 is the approximation to the \( \sin(\theta) \) function.

![Figure 1.6: Quadratic polynomial approximation of \( \sin(\theta) \).](image)

1.7 Error of Approximation

The general idea in this section is to determine, in the absence of the function being approximated being available, how well the polynomial will approximate it. The argument is rather convoluted in that we need to build a set of concepts and then bring them all together at the end, so please be patient.

**Step 1 (Define a function \( F(x) \))**

We start by defining

\[
F(x) = \prod_{i=0}^{n} (x - x_i) .
\]

For example for the special case of \( n = 2 \) we have the function

\[
F(x) = \prod_{i=0}^{2} (x - x_i) = (x - x_0)(x - x_1)(x - x_2)
\]

which is presented in Fig. 1.7.

It is important in this development to note that

\[
F(x) = \prod_{i} = 0^n (x - x_i) = 0 \text{ for } x = x_i, \quad i = 0, 1, \ldots, n
\]

because when the term \( (x_i - x_i) \) arises, it is zero and that eliminates the series. In other words, the function \( F(x) \) vanishes at the nodes.
10

CHAPTER 1. INTERPOLATION

Figure 1.7: Function $F(x)$ with zeros at $x_0, x_1,$ and $x_2$. The locations indicated by an $x$ along the axis are where $\frac{dF(x)}{dx} = 0$.

Because the approximation must equal the function at the nodal points $x_i$, we have

$$\hat{f}(x_i) = f(x_i) \quad i = 0, 1, ..., n. \quad (1.41)$$

Alternatively, we can write

$$f(x_i) - \hat{f}(x_i) = 0 \quad i = 0, 1, ..., n. \quad (1.42)$$

Step 2 (Define a function $g(x)$)

Now we change course and define the following function:

$$g(x) = f(x) - \hat{f}(x) - AF(x) \quad (1.43)$$

where $A$ is a constant. Why we do this has no answer at this point, but its relevance becomes evident shortly. Note from Eq. (1.40) and the Fig. 1.7 that there are $n + 1$ points $x_i$ where $F(x)$ is zero. Therefore, at these points, according to Eq. (1.42) and (1.40)

$$g(x_i) = 0 \quad i = 0, 1, ..., n. \quad (1.44)$$

We now digress once again in Step 3 which follows:

Step 3 (Introduction of the concept of $\xi$)

Let us choose $A$ such that $g(x) = 0$ at some arbitrary point $x_p \in [x_0, x_n]$. Now $g(x)$ is zero at least $n + 2$ points, that is at $x_0, x_1, ..., x_p, x_{n-1}, x_n$ (see Fig. 1.8). Since $g(x)$ is smooth, it must have a minimum or maximum between each pair of zeroes at which points the derivatives of $g(x)$ vanish. These are indicated by the letter $x$ in Fig. 1.8. Then $\frac{d(g)}{dx}$ has at least $(n + 2) - 1 = n + 1$ zeros in the interval $[x_0, x_n]$. Similarly at the points of inflection $\frac{d^2g(x)}{dx^2}$ has at least $(n + 2) - 1 = n$ zeros. Using similar logic to look at even higher derivatives, we finally arrive at the observation that $\frac{d^{n+1}g(x)}{dx^{n+1}}$ has $(n + 2) - (n + 1) = 1$ zero.

Let $x = \xi$ represent the location of this zero, that is

$$\left. \frac{d^{n+1}g(x)}{dx^{n+1}} \right|_{x = \xi} = 0 \quad (1.45)$$

Since $\hat{f}(x)$ is a polynomial of degree $n$, $\frac{d^{n+1}\hat{f}(x)}{dx^{n+1}} = 0$. This observation will be used a little later.
1.7. ERROR OF APPROXIMATION

Figure 1.8: Function $g(x)$ as presented in Eq. 1.43.

Step 4 (Calculation of $\frac{d^{n+1}F(x)}{dx^{n+1}}$)

It is not obvious, but true, that by differentiating

$$F(x) = \Pi_{i=0}^{n} (x - x_i)$$  \hspace{1cm} (1.46)

we obtain

$$\frac{d^{n+1}F}{dx^{n+1}} = (n+1)!$$  \hspace{1cm} (1.47)

To show this to be the case in at least one situation we provide the following example. This is not a proof, but provides some degree of confidence.

Example  Consider the example of $n = 1$

$$F(x) = (x - x_0)(x - x_1)$$  \hspace{1cm} (1.48)

$$\frac{dF(x)}{dx} = (x - x_0) + (x - x_1)$$  \hspace{1cm} (1.49)

$$\frac{d^2F(x)}{dx^2} = 1 + 1 = 2 = 2!$$  \hspace{1cm} (1.50)

Step 5 (Calculation of $A$)

Let us now differentiate $g(x)$ $n + 1$ times (see Eq. (1.43)) to give

$$\frac{d^{n+1}g(x)}{dx^{n+1}} = \frac{d^{n+1}f(x)}{dx^{n+1}} - \frac{d^{n+1}f^\prime(x)}{dx^{n+1}} - A \frac{d^{n+1}F(x)}{dx^{n+1}}.$$  \hspace{1cm} (1.51)

Consider what is happening at $x = \xi$ in Eq. (1.51) (we now add letters to refer to the terms):

$$\left.\frac{d^{n+1}g(x)}{dx^{n+1}}\right|_{x=\xi} = \xi - \left.\frac{d^{n+1}f(x)}{dx^{n+1}}\right|_{x=\xi} - \left.\frac{d^{n+1}f^\prime(x)}{dx^{n+1}}\right|_{x=\xi} - A \left.\frac{d^{n+1}F(x)}{dx^{n+1}}\right|_{x=\xi} = \xi$$  \hspace{1cm} (1.52)

We now address each of the terms in Eq. (1.52) one at a time. Term $D$ is zero because, from Eq. (1.45)

$$\left.\frac{d^{n+1}g(x)}{dx^{n+1}}\right|_{x=\xi} = 0.$$  \hspace{1cm} (1.53)
Term B is zero because we are taking the \( n+1 \) derivative of an \( n \)th degree polynomial.

\[
\frac{d^{n+1} \hat{f}(x)}{dx^{n+1}} \bigg|_{x=\xi} = 0
\]

From Eq. (1.47) we know that

\[
\frac{d^{n+1} F(x)}{dx^{n+1}} \bigg|_{x=\xi} = (n+1)!
\]

We now combine this information with Eq. (1.52) to give

\[
\frac{d^{n+1} f}{dx^{n+1}} \bigg|_{x=\xi} = A \frac{d^{n+1} F}{dx^{n+1}} \bigg|_{x=\xi} = A (n+1)!
\]

or, solving for \( A \)

\[
A = \frac{1}{(n+1)!} \frac{d^{n+1} f(x)}{dx^{n+1}} \bigg|_{x=\xi}
\]

Remember that \( A \) is chosen in such a way that \( g(x_p) = 0 \).

**Step 6 (Calculation of \( E(x) \))**

Now we know by the definition of \( x_p \), that at \( x = x_p \), \( g(x) = 0 \). Thus we have

\[
g(x_p) = f(x_p) - \hat{f}(x_p) - AF(x_p) = 0
\]

Rearranging Eq. (1.58) and substituting for \( A \) we obtain

\[
\hat{f}(x_p) - \hat{f}(x_p) = \frac{1}{(n+1)!} \frac{d^{n+1} f(x)}{dx^{n+1}} \bigg|_{x=\xi} = \xi (F(x_p))
\]

Finally, because \( x_p \) was selected arbitrarily, we can replace it with \( x \) to obtain:

\[
f(x) - \hat{f}(x) = \frac{1}{(n+1)!} \frac{d^{n+1} f(x)}{dx^{n+1}} \bigg|_{x=\xi} (F(x))
\]

so we can write, using the definition of \( F(x) \) from Eq. (1.46)

\[
E(x) = f(x) - \hat{f}(x) = \frac{1}{(n+1)!} \frac{d^{n+1} f(x)}{dx^{n+1}} \bigg|_{x=\xi} = \xi (\Pi_i = 0^n (x - x_i)) \quad \xi \in [x_0, x_n].
\]

Take a closer look at Eq. (1.61). We see that the error in the interpolation is inversely proportional to a function of the number of nodes, that is the term \( \frac{1}{(n+1)!} \). It is directly proportional to the value of the derivative, that is term \( \frac{d^{n+1} f(x)}{dx^{n+1}} \bigg|_{x=\xi} \), as one examines the values along \( x \). Finally it is directly proportional to a function of the size of the distance between the location \( x \) and the nodal locations \( x_i \), \( i = 0, ..., n \); such that the smaller the spacing between nodes, the smaller the error. For a given domain length, this is related to the value of \( n \). So a large \( n \) yields a small distance between nodes and they work together to yield a smaller error.
1.7. ERROR OF APPROXIMATION

Example 1
To see how one might use the above concept of error by revisiting the case presented earlier of

\[ f(x) = \ln(x) \quad x \in [1, 2]. \]  

(1.62)

From Eq. (1.30) on page 7 we have the approximation of \( \ln(x) \) using a piecewise linear Lagrange polynomial given as

\[ \hat{f}(x) = -0.693 + 0.693x \]  

(1.63)

and for \( x = 1.25 \) we have the following computed values and computed error

\[ \hat{f}(1.25) = 0.173 \]  

(1.64)

\[ f(1.25) = 0.22 \]  

(1.65)

\[ E(1.25) = 0.047. \]  

(1.66)

Now let us calculate the theoretical error using Eq. (1.61). Substituting the values for \( x = 1.25 \) we have

\[ E(x) = f(x) - \hat{f}(x) = \frac{1}{(n+1)!} \left. \frac{d^{n+1}f(x)}{dx^{n+1}} \right|_{x=\xi} (\Pi_{i=0}^{n} (x - x_i)) \quad \xi \in [x_0, x_n] \]  

(1.67)

which upon substitution of \( f(x) = \ln(x) \) and \( n = 1 \) gives

\[ E(x) = f(x) - \hat{f}(x) = \frac{1}{(1+1)!} \left. \frac{d^{1+1}\ln(x)}{dx^{1+1}} \right|_{x=\xi} (\Pi_{i=0}^{1} (x - x_i)) \quad \xi \in [1, 2] \]  

(1.68)

which yields

\[ E(1.25) = \frac{1}{2!} \left. \frac{d^{2}\ln(x)}{dx^{2}} \right|_{x=\xi} (x - x_0) (x - x_1) \]  

(1.69)

\[ E(1.25) = \frac{1}{2!} \left. \frac{d^{2}\ln(x)}{dx^{2}} \right|_{x=\xi} (1.25 - 1)(1.25 - 2) \]

\[ = \frac{1}{2!} \left( -\frac{1}{\xi^2} \right) (1.25 - 1)(1.25 - 2) \]

\[ = \frac{0.094}{\xi^2}. \]  

(1.70)

So what do we do with this? Well we recognize that

\[ 1 \leq \xi \leq 2 \]  

(1.71)

so, using the limiting (upper and lower bound) values of \( \xi \) we have

\[ \frac{0.094}{(2)^2} \leq E(1.25) \leq 0.094. \]  

(1.72)

or

\[ 0.023 \leq E(1.25) \leq 0.094. \]

Note that the actual error according to Eq. (1.64) is \( E(1.25) = 0.047 \), which is within the bounds indicated by Eq. (1.72).
Example 2

Let us consider the quadratic polynomial approximation to the \( \ln(x) \) \( x \in [1, 2] \). We have

\[
\hat{f}(x) = -1.21 + 1.33x - 0.214x^2
\]

which evaluated at \( x = 1.25 \) gives

\[
\hat{f}(1.25) = 0.213 \\
f(1.25) = 0.22 \\
E(1.25) = 0.006.
\]

From the error expression we have

\[
E(x) = f(x) - \hat{f}(x) = \frac{1}{(n + 1)!} \frac{d^{n+1}f(x)}{dx^{n+1}} \bigg|_{x=\xi} \left( \Pi_{i=0}^{n} (x - x_i) \right) \quad \xi \in [x_0, x_n] \quad (1.77)
\]

which, upon substitution of the function \( f(x) = \ln(x) \) and using \( n = 2 \), yields

\[
E(x) = f(x) - \hat{f}(x) = \frac{1}{(2 + 1)!} \frac{d^{2+1}\ln(x)}{dx^{2+1}} \bigg|_{x=\xi} \left( \Pi_{i=0}^{2} (x - x_i) \right) \quad \xi \in [1, 2] \quad (1.78)
\]

or, simplifying,

\[
E(x) = f(x) - \hat{f}(x) = \frac{1}{3!} \frac{d^{3}\ln(x)}{dx^{3}} \bigg|_{x=\xi} ((x - x_0)(x - x_1)(x - x_2)). \quad (1.79)
\]

Now we select the point of interest, that is \( x = 1.25 \), and substitute it into Eq. (1.79) to get

\[
E(1.25) = \frac{1}{3!} \left( \frac{2}{\xi^3} \right) (1.25 - 1)(1.25 - 1.5)(1.25 - 2.0) \\
= \frac{2}{(3)(2)\xi^3} (0.0469) = \frac{0.0156}{\xi^3}. \quad (1.80)
\]

Now we need to determine a choice of \( \xi \). If we use the upper and lower bounds of the interval \( x \in [1, 2] \) and therefore substitute \( \xi = 1 \) and \( \xi = 2 \) in Eq. (1.80) we obtain

\[
0.00195 \leq E(1.25) \leq 0.0156. \quad (1.81)
\]

Since the true error is \( E(1.25) = 0.006 \), Eq. (1.81) shows that it lies within the computed interval.

1.8 Multiple Elements

To this point we have been dealing with one interval. We will now introduce some new notation to consider multiple intervals. We define the interval over which one polynomial is defined as an \textit{element}. That element may have any number of nodes
1.8. MULTIPLE ELEMENTS

depending upon the degree of the polynomial. The higher the degree of Lagrange
polynomial, the larger the number of nodes we would use per element. We have thus
far considered as high as a quadratic Lagrange polynomial which required three nodes
per element.

In numerical methods we find the use of one element rather uninteresting. Rather
we want to concatenate several elements, and use a low degree Lagrange polynomial for
each element. In this section we extend our earlier work to consider multiple elements.

Firstly, consider a two element approximation using linear Lagrange polynomials in
each element. Such an arrangement is found in Fig. 1.9. Notice that there is a node in
common at the beginning and end of all internal elements. The elements on the ends
of the domain of interest share only one internal node. To introduce and illustrate
the multiple-element concept, we will use once again the function \( f(x) = \ln(x) \) as our
example. The values of \( f(x) = \ln(x) \) at each node, that is \( \ln(x_j) \) where \( j = 1, 2, 3 \) are
used to define the linear approximations

\[
\hat{f}(x) = \hat{\ln}(x) = \sum_{j=1}^{2} \ell_j(x) \ln(x_j) \quad x \in [1, 1.5] \quad (1.82)
\]

\[
\hat{f}(x) = \hat{\ln}(x) = \sum_{j=2}^{3} \ell_j(x) \ln(x_j) \quad x \in [1.5, 2.0]. \quad (1.83)
\]

Note that the indices in the summation now refer to node numbers, for example \( x_1 \)
denotes the value of \( x \) and node 1 and that to simplify notation from this point forward,
we will drop the superscript on \( \ell_j \) for the linear Lagrange polynomial.

These two equations describe the piecewise linear approximations between the
nodes in the upper panel of Fig. 1.9. In panel two, we see the relationship between
this approximation and the function being approximated, that is \( \ln(x) \). To proceed we

Figure 1.9: Two linear elements and their approximation (solid lines) of the logarithmic
function (dashed line).

need to find a convenient way to relate information at the local level, for example that
is that associated with the element, to that of the global system, for example that is
associated with the assemblage of elements in which the original problem is defined.
To achieve this goal consider the information provided in Fig. (1.10) and Tables 1.3
and 1.4. Figure. 1.10 shows the relationship between the global and local coordinate
systems. In the local coordinate system, each element sees the world from the same perspective. In other words, an observer at a point in the local system sees only what is happening on the element on which he/she resides. He/she does not see beyond the nodes defining each end of the element. The same observer when associating his/her position with respect to the global system sees the entire domain of interest inclusive of all elements.

It is helpful to distinguish more clearly between the two types of coordinates. In this spirit let us define the local coordinate as $\chi$. On the left is always node 0 and coordinate $\chi_0$ and on the right is always node 1 and coordinate $\chi_1$. Thus $0 \leq \chi \leq 1$ for each and every element. Similarly the Lagrange polynomials are always represented as a function of $\chi$, that is we have $\ell_0(\chi)$ and $\ell_1(\chi)$ no matter which element we are in.

From the global perspective you see in Fig. 1.10 that the node numbers are increasing from left to right as are the coordinate values. Table 1.3 presents the relationship between nodal numbering at the local or element scale and numbering at the global scale. Similarly, Table 1.4 shows the relationship between the global and nodal coordinates.

Let us see how we can derive a relationship between the two coordinate systems. We know that in element 1 when $\chi = \chi_0$, $x = x_1$. We also know that at $\chi = \chi_1$, $x = x_2$. It is clear that $x$ is a linear function of $\chi$, so we will write

$$x = a + b\chi.$$  \hspace{1cm} (1.84)

From our earlier observations we have

$$x_1 = a + b\chi_0 \hspace{1cm} (1.85)$$

$$x_2 = a + b\chi_1 \hspace{1cm} (1.86)$$

or

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & \chi_0 \\ 1 & \chi_1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}. \hspace{1cm} (1.87)$$

Solving for $a$ and $b$ we have

$$\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 1 & \chi_0 \\ 1 & \chi_1 \end{bmatrix}^{-1} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \frac{1}{\chi_1 - \chi_0} \begin{bmatrix} \chi_1 & -\chi_0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}. \hspace{1cm} (1.88)$$

Substitution of Eq. (1.88) into Eq. 1.84 yields

$$x = \frac{\chi_1 x_1 - \chi_0 x_2}{\chi_1 - \chi_0} + \frac{x_2 - x_1}{\chi_1 - \chi_0} \chi$$

$$= \frac{\chi_1 - \chi}{\chi_1 - \chi_0} x_1 + \frac{\chi - \chi_0}{\chi_1 - \chi_0} x_2. \hspace{1cm} (1.89)$$

We have seen this structure before; it is an expansion using Lagrange polynomials, in this case defined in the coordinate system $\chi$. In other words, we can write

$$x = \ell_0(\chi) x_1 + \ell_1(\chi) x_2$$

$$= \sum_{j=0}^{1} \ell_j(\chi) x(\chi_j) \hspace{1cm} (1.90)$$
where
\[ \ell_0 (\chi) = \frac{\chi_1 - \chi}{\chi_1 - \chi_0} \] (1.91)
and
\[ \ell_1 (\chi) = \frac{\chi - \chi_0}{\chi_1 - \chi_0}. \] (1.92)

Thus we can see that we can move between the coordinate systems; that is we can
determine a value of \( x \) given a value of \( \chi \) if we know the nodal locations \( x (\chi_0) \) and \( x (\chi_1) \) which are \( x_1 \), and \( x_2 \), respectively in our example. If we were to rewrite Eq. (1.84) as
\[ \chi = a + bx \] (1.93)
we could show that for element 2 (\( e = 2 \)) in Fig. 1.10
\[ \chi = \ell_2^2 (x) \chi \bigg|_{x=x_2} + \ell_3^3 (x) \chi \bigg|_{x=x_3} \] (1.94)
where the element number is indicated by the superscript on the left hand side of the
variable. Equation (1.94) states that, given a value of \( x \), we can determine the value
of \( \chi \).

1.8.1 Example
Suppose we want to find the value of \( x \) at location \( \chi = 0.5 \) in element 2 in Fig. 1.10. Using Eq. (1.90) we have
\[ x = \ell_0 (\chi) x_2 + \ell_1 (\chi) x_3 \]
\[ x|_{\chi=0.5} = \ell_0 (0.5) x_2 + \ell_1 (0.5) x_3 \]
\[ = (0.5) x_2 + (0.5) x_3 \]
\[ = \frac{x_2 + x_3}{2}. \] (1.95)
If we assume \( x_3 = 2 \) and \( x_2 = 1.5 \) as is the case in our approximation of \( \ln(x) \) as shown
in Fig. 1.9 we obtain
\[ x|_{\chi=0.5} = \frac{x_2 + x_3}{2} \]
\[ = \frac{1.5 + 2}{2} \]
\[ = \frac{3.5}{2}. \] (1.96)

For multiple elements we have what is shown in Fig. 1.11.

Let us now consider the approximations of our logarithm in each element in Fig.
1.9. In terms of the local coordinate \( \chi \) it will be of the form (recall that in this context
the element being considered is represented by the superscript 1 or 2):
Figure 1.10: Two linear elements in the global \((x)\) and local \((\chi)\) coordinate systems.

Figure 1.11: Global and local basis functions (Lagrange Polynomials) for four-node problem.

\[
\hat{f}^1(\chi) = \sum_{j=0}^{1} \ell^1_j(\chi) f^1(x_j) = \ell^1_0(\chi) f^1(x_0) + \ell^1_1(\chi) f^1(x_1) \quad x \in [1, 1.5] \quad (1.97)
\]

\[
\hat{f}^2(\chi) = \sum_{j=0}^{1} \ell^2_j(\chi) f^2(x_j) = \ell^2_0(\chi) f^2(x_0) + \ell^2_1(\chi) f^2(x_1) \quad x \in [1.5, 2] \quad (1.98)
\]

To obtain our approximation to the logarithm, we substitute appropriate values for \(f(\chi_j)^k\) where \(k\) is here used as a generalization of that used above specifically as the element number. For example, in Fig.1.9 \(f(x_1) = f(x)\bigg|_{x=0} = f^1(\chi)|_{\chi=0} = \ln(1) = 0\). Substituting the appropriate values observed at the nodes we get

\[
\hat{f}^1(\chi) = \ell^1_0(\chi)(0) + \ell^1_1(\chi)(0.40) \quad x \in [1, 1.5] \quad (1.99)
\]

\[
\hat{f}^2(\chi) = \ell^2_0(\chi)(0.4) + \ell^2_1(\chi)(0.693) \quad x \in [1.5, 2.0] \quad (1.100)
\]
Table 1.3: Relationship between local and global node numbering.

<table>
<thead>
<tr>
<th>Element</th>
<th>Local Node</th>
<th>Global Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 1.4: Relationship between local and global coordinates

If we now substitute in the definition of the linear Lagrange polynomials we obtain

\[
\hat{f}^1(\chi) = \left( \frac{\chi - \chi_1}{\chi_1 - \chi_0} \right)^1 (0) + \left( \frac{\chi_0 - \chi_1}{\chi_0 - \chi} \right)^1 (0.40)
\]

\[
\hat{f}^1(\chi) = \left( \frac{1.0 - \chi}{1.0} \right)^1 (0) + \left( \frac{0 - \chi}{-1.0} \right)^1 (0.40)
\]

\[
\hat{f}^1(\chi) = 0.0 + 0.40\chi^1 \quad x \in [1, 1.5]
\]  \hspace{1cm} (1.101)

and

\[
\hat{f}^2(\chi) = \left( \frac{\chi - \chi_1}{\chi_1 - \chi_0} \right)^2 (0.4) + \left( \frac{\chi_0 - \chi_1}{\chi_0 - \chi} \right)^2 (0.693)
\]

\[
\hat{f}^2(\chi) = \left( \frac{1.0 - \chi}{1.0} \right)^2 (0.40) + \left( \frac{0 - \chi}{-1.0} \right)^2 (0.693)
\]

\[
\hat{f}^2(\chi) = 0.29\chi^2 + 0.40 \quad x \in [1.5, 2.0]
\]  \hspace{1cm} (1.102)

which gives us an interpolation of the logarithm using the local coordinate system. Note that the superscripts on the bracketed quantities in Eqs. (1.101) and (1.102) represent the elements and are not exponents. Now let us calculate the value of \(\ln(1.25)\). We obtain

\[
\hat{f}^1(\chi = 0.5) = 0.20 \hspace{1cm} (1.103)
\]

since \(\hat{f}^1(\chi)\) has been obtained. For the location 1.75 we need to use the approximation in element 2. Thus we have

\[
\hat{f}(\chi = 0.5) = \hat{f}^2(\chi = (0.29)(0.5) + (0.4) = 0.546. \hspace{1cm} (1.104)
\]

Tabulating the results, including the quadratic approximation we obtained in Eq. (1.73), we get
Note for future reference that Eqs. (1.97) and (1.98) could be written in matrix notation for an arbitrary element $e$ as

$$
\left[ \hat{f}(\chi) \right]^e = \left[ \ell_0(\chi) \quad \ell_1(\chi) \right]^e \left[ \begin{array}{c} f(\chi_0) \\ f(\chi_1) \end{array} \right]^e.
$$

### 1.9 Hermite Polynomials

In this section we investigate another form of polynomial that allows us to directly interpolate not only the function $f(x)$ but also its derivative $\frac{df(x)}{dx}$; this is the Hermite polynomial. The Hermite polynomial is one of a suite of polynomials that when concatenated make up a numeric function that has a prescribed degree of smoothness. The point of departure for establishing the form of these piecewise Hermite polynomial functions is the following expression:

$$
f(x) = \sum_{j=0}^{1} \left( h_j^0(x) f(x_j) + h_j^1(x) \left. \frac{df}{dx} \right|_{x_j} \right) + E(x)
$$

where the functions $h_j^0(x)$ and $h_j^1(x)$ are the Hermite polynomials. The superscripts here are identified with the two forms of the Hermite function, one, $h_j^0$ associated with the value of the function at the node $x_j$, that is, $f(x_j)$, and the other $h_j^1(x)$ associated with the derivative at the node, that is $\left. \frac{df}{dx} \right|_{x_j}$. There are four functions, two associated with each node in the element, and they are shown in Fig. 1.12. To determine the form of the functions $h_j^0(x)$ and $h_j^1(x)$ we require that, as in the case of the Lagrange polynomials, the approximating function exactly represent at the nodes the function being approximated. However, in addition to these constraints we require that the approximation of the derivatives of the function being approximated also be exact at the nodes. These sets of requirement, two at each of two nodes, yields:

$$
\hat{f}(x_0) = f(x_0) 
$$

$$
\hat{f}(x_1) = f(x_1) 
$$

$$
\left. \frac{df}{dx} \right|_{x_0} = \left. \frac{df}{dx} \right|_{x_0} 
$$

$$
\left. \frac{df}{dx} \right|_{x_1} = \left. \frac{df}{dx} \right|_{x_1}.
$$

We now expand Eq. (1.106) for each of the above conditions; that is we expand the approximation and then impose the requirements of Eqs. (1.107)-(1.110). The first line of the two associated with each approximate value, for example $\hat{f}(x_0)$, is the expansion and the second is the value the expansion must represent.
1.9. HERMITE POLYNOMIALS

\[ f(x_0) = f(x_0) \]
\[ f(x_1) = f(x_1) \]

| \[ h_0^0 \] | \[ 1 \] | \[ 0 \] | \[ 0 \] | \[ 0 \] |
| \[ h_1^0 \] | \[ 0 \] | \[ 1 \] | \[ 0 \] | \[ 0 \] |
| \[ h_0^1 \] | \[ 0 \] | \[ 0 \] | \[ 0 \] | \[ 0 \] |
| \[ h_1^1 \] | \[ 0 \] | \[ 0 \] | \[ 0 \] | \[ 0 \] |

Table 1.5: Necessary conditions to be imposed on the Hermite polynomials to assure that the approximation and function are the same at the nodes.

| \[ \frac{df}{dx} \big|_{x_0} \] | \[ \frac{df}{dx} \big|_{x_0} \] | \[ \frac{df}{dx} \big|_{x_1} \] | \[ \frac{df}{dx} \big|_{x_1} \] |
| \[ dh_0^0 \big/ \big/ dx \big/ \big/ \] | \[ 0 \] | \[ 0 \] | \[ 0 \] |
| \[ dh_1^0 \big/ \big/ dx \big/ \big/ \] | \[ 0 \] | \[ 0 \] | \[ 0 \] |
| \[ dh_0^1 \big/ \big/ dx \big/ \big/ \] | \[ 1 \] | \[ 0 \] | \[ 0 \] |
| \[ dh_1^1 \big/ \big/ dx \big/ \big/ \] | \[ 0 \] | \[ 1 \] | \[ 0 \] |

Table 1.6: Necessary conditions to be imposed on the Hermite polynomials to assure that the approximation of the derivative and the derivative are the same at the nodes.

\[ \hat{f}(x_0) = \left[ \frac{h_0^0(x_0) f(x_0)}{f(x_0)} + h_0^1(x_0) f(x_1) + h_1^0(x_0) \frac{df}{dx} \big|_{x_0}\ + h_1^1(x_0) \frac{df}{dx} \big|_{x_1} \right] \]
\[ \hat{f}(x_1) = h_0^0(x_1) f(x_0) + \left[ \frac{h_1^0(x_1) f(x_1)}{f(x_1)} + h_0^1(x_1) \frac{df}{dx} \big|_{x_0}\ + h_1^1(x_1) \frac{df}{dx} \big|_{x_1} \right] \]
\[ \frac{df}{dx} \big|_{x_0} = \left[ \frac{dh_0^0}{dx} \big|_{x_0} f(x_0) + \frac{dh_1^0}{dx} \big|_{x_0} f(x_1) + \frac{dh_1^1}{dx} \big|_{x_0} \frac{df}{dx} \big|_{x_0}\ + \frac{dh_1^1}{dx} \big|_{x_1} \frac{df}{dx} \big|_{x_1} \right] \]
\[ \frac{df}{dx} \big|_{x_1} = \left[ \frac{dh_0^0}{dx} \big|_{x_1} f(x_0) + \frac{dh_1^0}{dx} \big|_{x_1} f(x_1) + \frac{dh_1^1}{dx} \big|_{x_0} \frac{df}{dx} \big|_{x_0}\ + \frac{dh_1^1}{dx} \big|_{x_1} \frac{df}{dx} \big|_{x_1} \right] \]

In general, for Eqs. (1.111) through (1.114) to be satisfied, for any one of the four equations, the boxed terms in each expression should be equal. For this to be true, the information appearing in the following tables is required. The first line of Table 1.5, for example, should be read as follows: In order to have \( \hat{f}(x_0) = f(x_0) \), we should have \( h_0^0 = 1 \). In addition, as seen in Table 1.6 \( h_0^0(x_0) \) and \( h_1^0(x_0) \) must also be zero at both \( x_0 \) and \( x_1 \) and \( h_1^0(x_0) = 0 \), where the locations \( x_0 \) and \( x_1 \) are the two nodal locations in the element.

From the first row in Tables 1.5 and 1.6 we see that there are four conditions to be imposed on \( h_0^0(x) \) (read horizontally across in each of these tables). For example,
$h_0^0(x)$ must take on a value of 1 at $x_0$ and 0 at $x_1$ and $\frac{dh_0^0(x)}{dx}$ must be 0 at both $x_0$ and $x_1$. Since a cubic polynomial is completely defined by four conditions (which is the number we have available for each of the functions $h_0^0(x)$, $h_1^0(x)$, $h_0^1(x)$, and $h_1^1(x)$), let us assume that each is a cubic polynomial of the form, for example:

$$h_0^0(x) = a_0 + b_0x + c_0x^2 + d_0x^3.$$ \hfill (1.115)

The coefficients $a_0, b_0, c_0,$ and $d_0$ can be determined using the information in the above tables. For example one equation for the case of $h_0^0(x)$ we can obtain using the first element of the first row in Table 1.5; that is $h_0^0(x)$ must be 1 at $x_0$. We get:

$$h_0^0(x_0) = a_0 + b_0x_0 + c_0x_0^2 + d_0x_0^3 = 1. \hfill (1.116)$$

If we now take the derivative of Eq. (1.115) and evaluate it at $x_0$ we know it must equal zero from the first row in Table 1.6. If we now collect this information for each element in row 1 of these tables we obtain the following set of equations for the coefficients $a_0, b_0, c_0,$ and $d_0$

$$\begin{bmatrix}
1 & x_0 & x_0^2 & x_0^3 \\
0 & 1 & 2x_0 & 3x_0^2 \\
1 & x_1 & x_1^2 & x_1^3 \\
0 & 1 & 2x_1 & 3x_1^2 \\
\end{bmatrix}
\begin{bmatrix}
a_0 \\
b_0 \\
c_0 \\
d_0 \\
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
0 \\
0 \\
0 \\
\end{bmatrix}
\begin{align}
h_0^0(x_0) &= 1 \\
\frac{dh_0^0(x_0)}{dx} &= 0 \\
h_0^0(x_1) &= 0 \\
\frac{dh_0^0(x_1)}{dx} &= 0. \hfill (1.117)
\end{align}$$

In reading Eq. (1.117) the information appearing to the right of the matrix equation identifies the conditions giving rise to each row of the equation.

Solving for $a_j, b_j, c_j,$ and $d_j$ and substituting in cubic polynomial we obtain

$$h_j^0(x) = \left(\ell_j^1(x)\right)^2 \left(1 - 2(x - x_j) \frac{d\ell_j^1(x)}{dx} \bigg|_{x_j}\right) \hfill (1.118)$$

where the Lagrange polynomial approximations are used in the definition. One can make similar arguments to obtain

$$h_j^1(x) = \left(\ell_j^1(x)\right)^2 (x - x_j). \hfill (1.119)$$

As noted earlier, the resulting functions are provided in Fig. 1.12. Note that the curves in this figure satisfy the constraints provided in Tables 1.5 and 1.6.
Example

Consider the function

\[ f(x) = \exp(x) \quad x \in [0.5, 1] \]  

(1.120)

which upon differentiation yields

\[ \frac{df(x)}{dx} = \exp(x) \quad x \in [0.5, 1]. \]  

(1.121)

The approximation \( \hat{f}(x) \) is given by introducing the values in Eq. (1.120) and (1.121) into Eq. (1.106); that is

\[ \hat{f}(x) = h_0^0(x) \exp(0.5) + h_1^0(x) \exp(0.5) + h_0^1(x) \exp(1.0) + h_1^1(x) \exp(1.0) \]  

(1.122)

A visual representation of the resulting approximation is provided in Fig. 1.13

![Hermite approximation of the exponential function.

Figure 1.13: Hermite approximation of the exponential function.

1.10 Error in Approximation by Hermites

We are not going to develop the error of the approximation using Hermites, but simply provide it for the our case (cubic with two nodes) below, viz. 


e^{x} = \sum_{n=0}^{\infty} \frac{x^{n}}{n!}.

Example of Hermite Error Approximation

\[ E(0.6) = \frac{1}{4!} \left[ (x - x_0)(x - x_1) \right]^2 \exp(x) \big|_{x=x} x \in [0.5, 1.0] \]
\[ = \frac{1}{4!} \left[ (0.6 - 0.5)(0.6 - 1.0) \right]^2 \exp(x) \big|_{x=x} \]
\[ = 6.67 \times 10^{-5} \exp(x) \big|_{x=x}. \]  

Since

0.5 \leq \xi \leq 1.0

then using these limits the error bound is

\[ 1.1 \times 10^{-4} \leq E(0.6) \leq 1.81 \times 10^{-4}. \]

The measured error is

\[ E(0.6) = 1.82212 - 1.82198 = 1.4 \times 10^{-4} \]

which lies within the error bounds calculated.

1.11 Chapter Summary

Polynomial approximation theory, the subject of this chapter, is a fundamental theoretical underpinning for numerical methods. In this chapter we introduce this topic in the context of using polynomial approximation theory to approximate functions. The Lagrangian polynomials and the Hermite polynomials are considered along with the error associated with using them in approximating functions. The concept of discretizing a domain, say a length along \( x \), into a set of concatenated linear segments called elements is introduced.

1.12 Problems

1. Determine (derive) the form of \( \ell_j^1(x) \) (see Fig. 1.14). The result should be an algebraic equation that describes the function \( \ell_j^1(x) \) for any interval \( (x_0, x_1) \). The strategy is to write the general form of the linear equation and then impose the conditions required of a linear approximating function.

2. Using Eq. (1.128) below write the form of the second degree or quadratic Lagrange polynomial \( \ell_j^2(x) \) shown below? How does it satisfy the requirements of a Lagrange polynomial?

\[ \ell_j^{(n)}(x) = \prod_{i=0}^{n} x - x_i \bigg|_{i \neq j} x_j - x_i. \]  

The shape of this function is shown in Fig. 1.15 below.
3. \( P_n(\theta) \) is a polynomial of degree \( n \), that is,

\[
P_n(\theta) = \sum_{i=0}^{n} a_i \theta^i
\tag{1.129}
\]

\( P_n(\theta_i) = f(\theta_i) \) where \( \theta_i \) are particular values of \( \theta \). Assume \( f(x) = \sin(\theta) \) (see Fig. 1.16) and that \( n = 3 \) (the polynomial is a cubic). Determine the polynomial represented by 1.129 using points \( \theta_i = 0, \pi/4, 3\pi/4, \pi \). Compare these errors to those presented in the example in the notes in Section 1.3. Sin (\( \theta \)) curve with measured points indicated by small black dots and interpolated values indicated by large black dots.

4. Consider the function

\[
f(x) = e^x \quad x \in [0, 1].
\tag{1.130}
\]

Use a piecewise linear polynomial to determine \( e^{0.5} \). Now calculate the error of your approximation, that is, determine the difference between the exact value of
$e^{0.5}$ and your estimate. Using the following equation for the theoretical error

$$E(x) = f(x) - \hat{f}(x) = \frac{1}{(n+1)!} \left. \frac{d^{n+1} f(x)}{dx^{n+1}} \right|_{x=\xi} (\prod_{i=0}^{n} (x - x_i)) \quad \xi \in [x_0, x_n]$$

(1.131)

Show that the error of your estimate is consistent with the theoretical estimate.

5. Consider the situation presented in Fig.1.17. You know the value of $x$ at the location at 1.5 in inches and you want to determine the same location in centimeters. Using the transformations given by

$$x = a + b\chi \quad (1.132)$$

where $x$ is the distance in inches and $\chi$ is the distance in centimeters determine the location $X$ in centimeters.

6 Please do the following:

(a) Using the information in Fig. 1.18, draw the linear Lagrange functions $\ell_i(x)$ and $\ell_i(\chi)$. In the local coordinate system use a solid line for $\ell_0(\chi)$ and a dashed line for $\ell_1(\chi)$. Label the horizontal and vertical axes with appropriate values. Give each of the symbolic $x_i$ and $\chi_i$ actual values, (in the sense of values of $x_0$ and $\chi_0$ etc.) and actual numerical values (in the sense of 1.0 etc.) in the spaces provided. You will have five values of $x_i$ of your choice and the same number of values of $\chi$, but there will be duplicate values of $\chi$ at most of the nodes.
1.12. PROBLEMS

(b) What is the value of $h = x_{i+1} - x_i$?
(c) What is the value of $\chi_1 - \chi_0$?

![Figure 1.18: Representation of five points in global ($x$) and local ($\chi$) coordinate systems.]

7. Given the formula for the Lagrange polynomial, that is

$$\ell^n_j (x) = \prod_{i=0}^{n} \frac{x - x_i}{x_j - x_i}$$

write the form of the cubic Lagrange polynomial. Then use this expression to write the approximation of the $\sin (\theta)$ segment from 0 to $\pi$. Use four equally spaced values of $\pi$ to obtain this approximation (four $\pi$ values). Now evaluate your polynomial at the points indicated in Table 1.1, to compute the error, and note how the errors compare with those already recorded in this table.

8. From the equation

$$x = \sum_{j=0}^{1} \ell_j (\chi) x (\chi_j) \quad \chi \in (0, 1), \quad x \in (1, 2)$$

(determine the value of $\chi$ at $x = 0.75$).

9. State the four requirements (constraints) necessary to obtain a Hermite interpolating polynomial $h_0^0$ as used in the expression

$$\hat{u} (x) = u_0 h_0^0 + u_1 h_1^0 + \frac{du}{dx} \bigg|_{x=0} h_0^1 + \frac{du}{dx} \bigg|_{x=1} h_1^1$$

(1.135)

assuming you start with the standard cubic polynomial.

10. Use a linear polynomial to approximate $\sin \theta$, $0 \leq \theta \leq \pi$. Shown in Fig. 1.19. Determine whether the error at $\pi/2$ falls within the range predicted by the error formula.

11. Use quartic (fourth degree) Lagrange polynomials to approximate $\sin \theta$ using nodes at $0, \pi/2, \pi, (3/2) \pi$ and $\pi$. Determine the value of $\sin \theta$ at $(5/4) \pi$. Compare the estimate to the value obtained using the quadratic Lagrange polynomial. Is it better; why (or why not)?
12. Given the \( \sin(\theta) \) function shown above and linear polynomials written in terms of the local coordinate system \( \chi \), determine the value of the \( \sin(\theta) \) at \( \chi = 0.5 \).

13. The goal here is to determine if one can use any two points along the line from \( x = 0 \) to \( x = 2 \) in Fig. 1.20 to estimate the value at \( x = 1.0 \). Please solve this problem by following these steps:

(a) Assume \( f(x) \) can be approximated by

\[
\hat{f}(x) = \sum_{j=0}^{1} \ell_j(x) f_j
\]

where \( \ell_j(x) \) is the linear Lagrange polynomial and \( f_j \) is the value of \( f(x) \) at \( x = x_j \).

(b) Expand this expression to represent that needed for this problem.

(c) Evaluate the resulting expression at the point \( x = 0.0 \) and \( x = 0.5 \). Assume \( \hat{f}(0.0) = 0.25 \) and \( \hat{f}(0.5) = 0.5 \). You should have two equations in the two unknown values of \( f_1 \) and \( f_2 \) where the subscripts are the two nodal identifiers, that is nodes 1 and 2.

(d) Solve for the two unknown values of \( f_1 \) and \( f_2 \).

(e) Use these values to estimate the value of \( f(x) \), where \( x = 1.0 \).

Hint: Remember that the linear Lagrange polynomials are 1 at the node for which they are defined and 0 elsewhere, so in this case they are two straight lines crossing at \( x = 1 \) and \( \ell(1) = 0.5 \).
1.12. PROBLEMS

Figure 1.20: Problem definition.
Bibliography


Chapter 2

Numerical Differentiation

2.1 General Theory

We are now ready to use some of the theory we developed above to derive the approximation of a continuous derivative that will yield an algebraic approximate equation that can be used to solve a differential equation numerically using elementary algebra. Let us begin with the following interpolation formula where \( \ell^n_j(x) \) are the Lagrange polynomials of degree \( n \)

\[
 f(x) = \sum_{j=0}^{n} \ell^n_j(x) f(x_j) + E(x).
\]

Now differentiate \( f(x) \) to give

\[
 \frac{df(x)}{dx} = \sum_{j=0}^{n} \frac{d\ell^n_j(x)}{dx} f(x_j) + \frac{dE(x)}{dx}.
\]

Notice that in differentiating \( f(x) \) the value of the function at location \( x_j \) is not affected because it is a number, not a function of \( x \).

From the equation defining the error associated with approximating functions using Lagrange polynomials, we have from Eq. (1.61) on page 12

\[
 E(x) = \frac{1}{(n+1)!} \Pi^n_{i=0} (x - x_j) \left[ \frac{d^{(n+1)}f(x)}{dx^{(n+1)}} \right]_{x=\xi}.
\]

Again, differentiating we have

\[
 \frac{dE(x)}{dx} = \frac{d}{dx} \left[ \frac{1}{(n+1)!} \Pi^n_{i=0} (x - x_j) \left[ \frac{d^{(n+1)}f(x)}{dx^{(n+1)}} \right]_{x=\xi} \right]
\]

where we must remember that \( \xi \) is an unknown function of \( x \). In order to proceed, we need to introduce the results of a theorem that will help us understand the term on the right hand side of Eq. (2.4). A proof of this theorem is found in Ralston, (1965, pg. 77).

**Theorem 2** Let

\[
 E(x) = \frac{1}{(n+1)!} \Pi^n_{i=0} (x - x_j) \left[ \frac{d^{(n+1)}f(x)}{dx^{(n+1)}} \right]_{x=\xi}
\]
be the error term in the Lagrangian interpolation formula with $\xi$ in the interval spanned by $x_0, x_1, \ldots, x_n$ and $x$. Then if $\frac{d^{(n+1)}f(x)}{dx^{(n+1)}}$ is continuous,

$$
\frac{1}{(n+1)!} \frac{d}{dx} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \right) \bigg|_{x=\xi} = \frac{1}{(n+2)!} \frac{d^{(n+2)} f(x)}{dx^{(n+2)}} \bigg|_{x=\eta}
$$

(2.6)

where $\eta$ is also in the interval spanned by $x_0, x_1, \ldots, x_n$ where $x_i = i h$ and

$$
h \equiv x_i - x_{i-1}.
$$

(2.7)

Combination of Eq. (2.4) with the results of the above theorem give

$$
\frac{dE(x)}{dx} = \frac{d}{dx} \left[ \frac{1}{(n+1)!} \Pi_{i=0}^{n} (x - x_i) \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \bigg|_{x=\xi} \right]
$$

$$
= \Pi_{i=0}^{n} (x - x_i) \frac{1}{(n+2)!} \frac{d^{(n+2)} f(x)}{dx^{(n+2)}} \bigg|_{x=\eta}
$$

$$
+ \frac{1}{(n+1)!} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \bigg|_{x=\xi} \right) \frac{d}{dx} \left[ \Pi_{i=0}^{n} (x - x_i) \right].
$$

(2.8)

Note that at a nodal point $x_j$ the first term on the right-hand side of Eq. (2.8) vanishes (because we are forming a product and when $x$ is a nodal location, one term in the product is zero, so the whole product is zero) and we have

$$
\frac{dE}{dx} \bigg|_{x_j} = \frac{1}{(n+1)!} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \bigg|_{x=\xi} \right) \frac{d}{dx} \left[ \Pi_{i=0}^{n} (x - x_i) \right]_{x_j}.
$$

(2.9)

Notice that we take the derivative $\frac{d}{dx} \left[ \Pi_{i=0}^{n} (x - x_i) \right]$ and then evaluate it at $x_j$. If we were to do the reverse, that is, if we introduced $x_j$ first then differentiated, we would be taking the derivative of a number and the term would vanish. Now we can write

$$
\frac{df}{dx} \bigg|_{x_k} = \sum_{j=0}^{n} \frac{df}{dx} \bigg|_{x_k} f(x_j) + \frac{1}{(n+1)!} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \bigg|_{x=\xi} \right) \frac{d}{dx} \left[ \Pi_{i=0}^{n} (x - x_i) \right]_{x_k}
$$

(2.10)

which says that the derivative $\frac{df(x)}{dx}$ at $x_k$ can be represented by the right-hand-side of Eq. (2.10).

### 2.2 Two-Point Difference Formulae

We are now in a position to look at some specific difference approximations arising out of the expression presented in Eq. (2.10). Consider

$$
\frac{df}{dx} \bigg|_{x_i} = \frac{df}{dx} \bigg|_{x_i} + \frac{dE}{dx} \bigg|_{x_i}.
$$

(2.11)

Let us take the particular representation of $\frac{df}{dx} \bigg|_{x_i}$ presented in the following equation:

$$
\frac{df}{dx} \bigg|_{x_i} = \sum_{j=0}^{n} \frac{df}{dx} \bigg|_{x_i} f(x_j).
$$

(2.12)
2.2. TWO-POINT DIFFERENCE FORMULAE

This equation should be read as the derivative of the approximation, that is \( \frac{df}{dx} \), evaluated at the location \( x_i \). We will look first at the case of using the linear Lagrange polynomial \( \ell_j^i (x) \). Expansion of Eq. (2.12) for the two terms associated with having \( n = 1 \) gives

\[
\frac{df}{dx} \bigg|_{x_i} = \frac{d\ell_0}{dx} \bigg|_{x_i} f(x_0) + \frac{d\ell_1}{dx} \bigg|_{x_i} f(x_1)
\]  

(2.13)

where the linear Lagrange polynomials are provided by the definitions

\[
\ell_0 (x) = \frac{x-x_1}{x_0-x_1}
\]  

(2.14)

and

\[
\ell_1 (x) = \frac{x-x_0}{x_1-x_0}.
\]  

(2.15)

The associated derivatives of these two functions are given by

\[
\frac{d\ell_0 (x)}{dx} = \frac{1}{x_0-x_1}
\]  

(2.16)

and

\[
\frac{d\ell_1 (x)}{dx} = \frac{1}{x_1-x_0}.
\]  

(2.17)

Combining Eqs. (2.16) and (2.17) with (2.13) we have

\[
\frac{df}{dx} \bigg|_{x_i} = \frac{1}{x_0-x_1} f(x_0) + \frac{1}{x_1-x_0} f(x_1)
\]  

(2.18)

or, rearranging

\[
\frac{df}{dx} \bigg|_{x_i} = \frac{f(x_1) - f(x_0)}{x_1-x_0}
\]  

(2.19)

or

\[
\frac{df}{dx} \bigg|_{x_i} = \frac{f(x_1) - f(x_0)}{h} \quad x_i = 0, 1.
\]  

(2.20)

We call this a finite difference approximation to \( \frac{df}{dx} \bigg|_{x_i} \). Note that we have not specified the value of \( i \), so this approximation can be envisioned as being identified with (located at) either \( i = 0 \) or \( i = 1 \). However, it is sometimes important to know where the approximation is perceived to be located, so the following nomenclature is used.

2.2.1 Forward Difference Formula

\[
\frac{df}{dx} \bigg|_{x_0} = \frac{f(x_1) - f(x_0)}{h} \equiv \frac{\Delta f}{h} \bigg|_{x_0}.
\]  

(2.21)

\[1\] We will drop the superscript designating the degree of the polynomial when it is clear by the context the degree that is intended.
2.2.2 Backward Difference Formula

\[ \frac{df}{dx} \bigg|_{x_1} = \frac{f(x_1) - f(x_0)}{h} \equiv \frac{\nabla f(x_1)}{h} \bigg|_{x_1}. \]  
(2.22)

Note that the symbol \( \nabla (\cdot) \) is not the gradient operator \( \nabla (\cdot) \) from the vector calculus which is a vector quantity.

2.2.3 Example

Consider differentiation of \( \ln(x) \) evaluated at location \( x_i \), that is

\[ \frac{df}{dx} \bigg|_{x_i} = \frac{d\ln(x)}{dx} \bigg|_{x_i} = \frac{1}{x_i} \quad x \in [1, 2]. \]  
(2.23)

We begin with the definition of our approximating derivative \( \frac{df}{dx} \)

\[ \frac{df}{dx} = \frac{df}{dx} + \frac{dE}{dx}. \]  
(2.24)

Neglecting for the moment the error term \( \frac{dE}{dx} \), we have from Eq. (2.19)

\[ \frac{df}{dx} \bigg|_{x_i} = \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \]  
(2.25)

Since, in this example, we have as the values of \( f(x_0) \) and \( f(x_1) \)

\[ f(x_0) = \ln(1) = 0 \]  
(2.26)

and

\[ f(x_1) = \ln(2) = 0.693 \]  
(2.27)

we get, through substitution in Eq. (2.25) the following:

\[ \frac{df}{dx} \bigg|_{x_i} = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = \frac{0.693 - 0}{2 - 1} = 0.693 \quad i = 0, 1. \]  
(2.28)

Let us now investigate the error of this approximation. The exact derivative \( \frac{df}{dx} \), our approximation \( \frac{df}{dx} \), and the error \( \frac{dE}{dx} \) are tabulated below:

<table>
<thead>
<tr>
<th>( x )</th>
<th>( \frac{df}{dx} )</th>
<th>( \frac{dE}{dx} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_0 )</td>
<td>1</td>
<td>0.693</td>
</tr>
<tr>
<td>( x_1 )</td>
<td>0.5</td>
<td>0.693</td>
</tr>
</tbody>
</table>

(2.29)

2.2.4 Error of the Approximation

In the above table we have documented the observed error in the derivative approximation when a forward or backward difference is used. Now we will look at how this relates to the error as described Eq. (2.9) on page 34 which we rewrite here for convenience as:

\[ \frac{dE}{dx} \bigg|_{x_j} = \frac{1}{(n+1)!} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \right) \bigg|_{x=x_j} \frac{d}{dx} \left[ \prod_{i=0}^{n} (x - x_i) \right] \bigg|_{x_j}. \]  
(2.30)
For linear elements this equation becomes

\[
\frac{dE}{dx} \bigg|_{x_j} = \frac{1}{(1 + 1)!} \left( \frac{d^{(1+1)} f (x)}{dx^{(1+1)}} \right) \bigg|_{x = \xi} \frac{d}{dx} \left[ \Pi_{i=0}^{1} (x - x_i) \right] \bigg|_{x_j}
\] (2.31)

or

\[
\frac{dE}{dx} \bigg|_{x_j} = \frac{1}{2!} \frac{d^2 f (x)}{dx^2} \bigg|_{x = \xi} \left[ (x_j - x_0) + (x_j - x_1) \right].
\] (2.32)

If we evaluate this at \( x_0 \) we have

\[
\frac{dE}{dx} \bigg|_{x_0} = \frac{1}{2!} \frac{d^2 f (x)}{dx^2} \bigg|_{x = \xi} (x_0 - x_1)
\]

\[
= -\frac{1}{2!} \frac{d^2 f (x)}{dx^2} \bigg|_{x = \xi} h
\]

\[
= O (h)
\] (2.33)

where \( O (h) \) is read as ‘order \( h \)’. Substitution of \( f (x) = \ln(x) \) gives

\[
\frac{dE}{dx} \bigg|_{x_0} = -\frac{1}{2!} \frac{d^2 P (x)}{dx^2} \bigg|_{x = \xi} h
\]

\[
= -\frac{1}{2!} \frac{1}{(-\xi^2)} h
\]

\[
= \frac{1}{2!} \frac{1}{\xi^2} h
\] (2.34)

If we introduce the limits \( \xi \in [1, 2] \) and remember that \( h = 1 \) because \( x \in [1, 2] \) and there are only two nodes, one at \( x_0 \) and the other at \( x_1 \), we have

\[
\left( \frac{1}{2!} \right) \left( \frac{1}{2^2} \right) (1) \leq \frac{dE}{dx} \bigg|_{x_0} \leq \left( \frac{1}{2!} \right) \left( \frac{1}{1^2} \right) (1)
\] (2.35)

or

\[
\frac{1}{8} \leq \frac{dE}{dx} \bigg|_{x_0} \leq \frac{1}{2}
\] (2.36)

or

\[
0.125 \leq \frac{dE}{dx} \bigg|_{x_0} \leq 0.5.
\] (2.37)

Thus our observed value of the error of 0.307 as shown in the above table lies within the predicted error bounds.

The bounds for \( \frac{dE}{dx} \bigg|_{x_1} \) can be calculated as

\[
-0.5 \leq \frac{dE}{dx} \bigg|_{x_1} \leq -0.125
\] (2.38)

so the computed error at \( x_1 \), namely -0.193 is also within the expected error bounds.

2.3 Two-Point Formulae from Taylor Series

In this section we present an alternative way of obtaining two-point difference formula. We begin with the definition of a Taylor series, viz.:
If \( f(x) \) has continuous derivatives of order \( k+1 \) in \([a, b]\) then by Taylor’s theorem

\[
f(x_j + h) = f(x_j) + h \frac{df}{dx} \bigg|_{x_j} + \cdots + \frac{h^k}{k!} \frac{d^k f}{dx^k} \bigg|_{x_j} + \frac{h^{(k+1)}}{(k+1)!} \frac{d^{(k+1)} f}{dx^{(k+1)}} \bigg|_{x_j + \theta_j h} \theta_j \in [0, 1]
\]

which we denote as a Taylor series.

Let us consider the interval

\[
x \in [x_0, x_1].
\]

Then from the Taylor series we have (keeping in mind that \( h = x_1 - x_0 \) and that we are looking backwards and therefore \( h \) is negative)

\[
f(x_1 - h) = f(x_1 - (x_1 - x_0)) = f(x_0)
\]

\[
= f(x_1) - h \frac{df}{dx} \bigg|_{x_1} + \frac{h^2}{2!} \frac{d^2 f}{dx^2} \bigg|_{x_1 + \theta_h} \theta \in [0, 1].
\]

Rearranging the terms in Eq. (2.41) we have:

\[
\frac{df}{dx} \bigg|_{x_1} = \frac{f(x_1) - f(x_0)}{h} + \frac{h}{2!} \frac{d^2 f}{dx^2} \bigg|_{x_1 + \theta_h} \theta \in [0, 1]
\]

where the labels under the terms indicate their relationship to our earlier development.

We now compare Eq. (2.42) with the following expressions developed previously using the Lagrange polynomial approach

\[
\frac{df}{dx} \bigg|_{x_i} = \frac{d\hat{f}}{dx} \bigg|_{x_i} + \frac{dE}{dx} \bigg|_{x_i}.
\]

Employing Eq. (2.25) we have

\[
\frac{d\hat{f}}{dx} \bigg|_{x_1} = \frac{f(x_1) - f(x_0)}{h}
\]

and from Eq. (2.32) we obtain

\[
\frac{dE}{dx} \bigg|_{x_1} = \frac{1}{2!} \frac{d^2 f}{dx^2} \bigg|_{x_1} (x_1 - x_0)
\]

\[
= \frac{1}{2!} \frac{d^2 f}{dx^2} \bigg|_{x_1} h.
\]

If we combine the above three equations we obtain

\[
\frac{df}{dx} \bigg|_{x_1} = \frac{f(x_1) - f(x_0)}{h} + \frac{1}{2!} \frac{d^2 f}{dx^2} \bigg|_{x_1} h.
\]
Equation (2.46) is equivalent to Eq. (2.42).

At this point we have shown how to develop an approximation of a first derivative using two equivalent strategies. In the following subsubsection we demonstrate how one can use such an approximation to solve a simple problem. This is an elementary application of the method of finite differences first introduced by the famous mathematician Euler in 1768.

**Example Solution of a First-Order Equation**

Consider the physical system illustrated in Fig. 2.1. Fluid is entering the pipe with cross-sectional area $A$ at a rate $Q$ per unit length $[L]$ and has units $L^2 T$ per unit length $[L]$. The velocity entering the pipe is $v = 3$ with units $L T$. What is the velocity of the pipe along its length? The governing equation, boundary condition, and known flux $Q$ are given as:

$$A \frac{dv}{dx} = Q, \quad x \in [0, 3]$$  \hspace{1cm} (2.47)

$$v(0) = 3$$  \hspace{1cm} (2.48)

$$Q = 1.$$  \hspace{1cm} (2.49)

The analytical solution is

$$v = Qx + a$$  \hspace{1cm} (2.50)

$$v(0) = Q(0) + a$$  \hspace{1cm} (2.51)

$$a = 3$$  \hspace{1cm} (2.52)

$$v = x + 3.$$  \hspace{1cm} (2.53)

![Figure 2.1: Example problem for application of a first derivative finite difference approximation.](image)

Now let us solve this physical problem using our numerical approximation. Employing a first-order finite difference representation for the derivative we get

$$\frac{dv}{dx} = Q$$  \hspace{1cm} (2.54)

$$\frac{dv}{dx}|_{x_0} = \frac{v(x_1) - v(x_0)}{x_1 - x_0} = \frac{v_1 - v_0}{1}. $$  \hspace{1cm} (2.55)

Using the same general formula for each nodal location we obtain
\[ \left( \frac{d\hat{v}}{dx} - Q \right) \bigg|_{x_0} = 0 \Rightarrow \frac{v_1 - v_0}{1} - 1 = 0 \] (2.56)

\[ \left( \frac{d\hat{v}}{dx} - Q \right) \bigg|_{x_1} = 0 \Rightarrow \frac{v_2 - v_1}{1} - 1 = 0 \] (2.57)

\[ \left( \frac{d\hat{v}}{dx} - Q \right) \bigg|_{x_2} = 0 \Rightarrow \frac{v_3 - v_2}{1} - 1 = 0. \] (2.58)

In matrix form we have

\[
\begin{bmatrix}
-1 & 1 & 0 & 0 \\
0 & -1 & 1 & 0 \\
0 & 0 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
v_0 \\
v_1 \\
v_2 \\
v_3
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}.
\] (2.59)

But we know from the boundary conditions that \( v(0) = 3 \). Thus we can write

\[
\begin{bmatrix}
-1 & 1 & 0 & 0 \\
0 & -1 & 1 & 0 \\
0 & 0 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
3 \\
v_1 \\
v_2 \\
v_3
\end{bmatrix}
= 
\begin{bmatrix}
1 \\
1 \\
1
\end{bmatrix}.
\] (2.60)

Since we know \( v_0 \) to be 3, we can multiply each coefficient in the first column by 3 and, after changing the sign, place the result on the right-hand side of the equation. The first column of the matrix can now be removed. The result follows:

\[
\begin{bmatrix}
1 & 0 & 0 \\
-1 & 1 & 0 \\
0 & -1 & 1
\end{bmatrix}
\begin{bmatrix}
v_1 \\
v_2 \\
v_3
\end{bmatrix}
= 
\begin{bmatrix}
1 + 3 \\
1 + 0 \\
1 + 0
\end{bmatrix}.
\] (2.61)

The values of the unknown velocities are now given by

\[
\begin{bmatrix}
v_1 \\
v_2 \\
v_3
\end{bmatrix}
= 
\begin{bmatrix}
1 & 0 & 0 \\
-1 & 1 & 0 \\
0 & -1 & 1
\end{bmatrix}^{-1}
\begin{bmatrix}
1 + 3 \\
1 + 0 \\
1 + 0
\end{bmatrix}
= 
\begin{bmatrix}
4 \\
5 \\
6
\end{bmatrix}.
\] (2.62)

The result obtained numerically is the same as we would get if we evaluated Eq. (2.53) for the nodal locations shown in Fig. 2.1. The reason it is exact is that the solution is a straight line and we are representing it with piecewise linear functions. From Eq. (2.32) on page 37 we see that when the solution is linear the error of the approximation vanishes.

### 2.4 Three-point Difference Formulae

Now we will extend the above concept to consider a three-node element. We will see that with three nodes we can get a more accurate approximation of the first derivative and an approximation of a second derivative (we could not get an approximation of a second derivative given linear Lagrange polynomials and one element with two nodes).

From the general statement expressing a first derivative in terms of a Lagrangian expansion, that is Eq. (2.2) on page 33, we have
2.4. THREE-POINT DIFFERENCE FORMULAE

\[ \frac{df}{dx} = \sum_{j=0}^{n} \frac{d\ell^j_n(x)}{dx} f(x_j) + \frac{dE(x)}{dx}. \]  

(2.63)

Now consider the second-degree Lagrange polynomial (which involves three nodes and \( n = 2 \)), that is

\[ \frac{df}{dx} = 2 \sum_{j=0}^{2} \frac{d\ell^j_2(x)}{dx} f(x_j) + \frac{dE(x)}{dx}. \]  

(2.64)

2.4.1 First-Order Derivative Difference Formulae

From Eq. (2.64) we have for the case of a quadratic Lagrange polynomial

\[ \frac{df}{dx} = \frac{d\ell_0(x)}{dx} f(x_0) + \frac{d\ell_1(x)}{dx} f(x_1) + \frac{d\ell_2(x)}{dx} f(x_2) + \frac{dE(x)}{dx} \]  

(2.65)

which upon substitution of the definitions of the quadratic Lagrange polynomials gives:

\[ \frac{df}{dx} = \left( \frac{2x - x_1 - x_2}{2h^2} \right) f(x_0) + \left( \frac{2x - x_0 - x_2}{-h^2} \right) f(x_1) \]

\[ + \left( \frac{2x - x_0 - x_1}{2h^2} \right) f(x_2) + \frac{dE(x)}{dx}. \]  

(2.66)

Let us evaluate this expression at node \( x_1 \). We obtain

\[ \frac{df}{dx} \bigg|_{x_1} = \left( \frac{2x_1 - x_1 - x_2}{2h^2} \right) f(x_0) + \left( \frac{2x_1 - x_0 - x_2}{-h^2} \right) f(x_1) \]

\[ + \left( \frac{2x_1 - x_0 - x_1}{2h^2} \right) f(x_2) + \frac{dE}{dx} \bigg|_{x_1}. \]

(2.67)

Notice that this formula is peculiar in that the node at which the derivative is being defined does not appear in the approximation.
Now let us consider the case for node \( x_0 \)
\[
\frac{df}{dx}\bigg|_{x_0} = \left( \frac{2x_0 - x_1 - x_2}{2h^2} \right) f(x_0) + \left( \frac{2x_0 - x_0 - x_2}{-h^2} \right) f(x_1)
\]
\[
+ \frac{(2x_0 - x_0 - x_1)}{2h^2} f(x_2) + \frac{dE}{dx}\bigg|_{x_0}
\]
\[
= \left( \frac{x_0 - x_1 + x_0 - x_2}{2h^2} \right) f(x_0) + \left( \frac{x_0 - x_2}{-h^2} \right) f(x_1) + \left( \frac{x_0 - x_1}{2h^2} \right) f(x_2) + \frac{dE}{dx}\bigg|_{x_0}
\]
\[
= \left( -\frac{h - 2h}{2h^2} \right) f(x_0) + \left( -\frac{2h}{-h^2} \right) f(x_1) + \frac{h}{2h^2} f(x_2) + \frac{dE}{dx}\bigg|_{x_0}
\]
\[
= \left( \frac{-3}{2h} \right) f(x_0) + \left( \frac{2}{h} \right) f(x_1) + \left( -\frac{1}{2h} \right) f(x_2) + \frac{dE}{dx}\bigg|_{x_0}
\]
\[
= \frac{1}{2h} \left( -3f(x_0) + 4f(x_1) - f(x_2) \right) + \frac{dE}{dx}\bigg|_{x_0}.
\]
(2.68)

Note that, based upon what we have provided above, it appears that when using three nodes and quadratic Lagrange polynomials, the derivative approximation has a different form depending on where it is being evaluated. In other words
\[
\frac{df}{dx}\bigg|_{x_0} \neq \frac{df}{dx}\bigg|_{x_1} \neq \frac{df}{dx}\bigg|_{x_2}.
\]
(2.69)

**Error of the Approximation**

The general definition of the error in representing the first derivative is given by
\[
\frac{dE}{dx} = \frac{1}{(n + 1)!} \left( \frac{d^{n+1}f(x)}{dx^{n+1}} \right)_{x=x_\xi} \frac{d}{dx} \left[ \prod_{i=0}^{n} (x - x_i) \right]
\]
(2.70)

which, for quadratic Lagrange polynomials, gives for node \( x_1 \)
\[
\frac{dE}{dx}\bigg|_{x_1} = \frac{1}{3!} \left( \frac{d^{(3)}f(x)}{dx^{(3)}} \right)_{x=x_\xi}
\]
\[
\left[ \frac{(x_1 - x_0)(x_1 - x_1)}{0} + \frac{(x_1 - x_1)(x_1 - x_2)}{0} + \frac{(x_1 - x_0)(x_1 - x_2)}{-h^2} \right]
\]
(2.71)

\[
\frac{dE}{dx}\bigg|_{x_1} = \frac{-1}{3!} \left( \frac{d^{(3)}f(x)}{dx^{(3)}} \right)_{x=x_\xi} h^2
\]
\[
= O(h^2).
\]
(2.72)

Note that this \( O(h^2) \) compares with \( O(h) \) for the linear polynomial approximation of the first derivative at node \( x_1 \) (see Eq. (2.33) on page 37), that is
\[
\frac{dE}{dx}\bigg|_{x_1} = \frac{1}{(1 + 1)!} \frac{d^2f(x)}{dx^2} \bigg|_{x=x_\xi}(x_1 - x_0)
\]
\[
= \frac{1}{2!} \frac{d^2f(x)}{dx^2} \bigg|_{x=x_\xi} h
\]
\[
= O(h).
\]
(2.73)
2.4. TWO-POINT DIFFERENCE FORMULAE

2.4.2 Second-Order Derivatives

To get an approximation for the second-order derivative of \( f(x) \), differentiate Eq. (2.64) to give

\[
\frac{df^2(x)}{dx^2} = \sum_{j=0}^{3} \frac{d^2 \ell_j^2(x)}{dx^2} f(x_j) + \frac{d^2 E(x)}{dx^2}.
\]  

(2.74)

Expanding this equation we obtain

\[
\frac{df^2(x)}{dx^2} = \frac{d^2 \ell_0^2(x)}{dx^2} f(x_0) + \frac{d^2 \ell_1^2(x)}{dx^2} f(x_1) + \frac{d^2 \ell_2^2(x)}{dx^2} f(x_2) + \frac{d^2 E(x)}{dx^2}.
\]  

(2.75)

Substitution of the definitions of \( \ell_j^2(x) \) yields

\[
\frac{df^2(x)}{dx^2} = \frac{d}{dx} \left( \frac{2x - x_1 - x_2}{2h^2} \right) f(x_0) + \frac{d}{dx} \left( \frac{2x - x_0 - x_2}{-h^2} \right) f(x_1)
\]
\[
+ \frac{d}{dx} \left( \frac{2x - x_0 - x_1}{2h^2} \right) f(x_2) + \frac{d^2 E(x)}{dx^2}.
\]  

(2.76)

Differentiating we obtain

\[
\frac{df^2(x)}{dx^2} = \frac{1}{h^2} f(x_0) - \frac{2}{h^2} f(x_1) + \frac{1}{h^2} f(x_2) + \frac{d^2 E(x)}{dx^2}.
\]  

(2.77)

If we neglect the error term we can now write the approximated derivative as

\[
\frac{df^2(x)}{dx^2} = \frac{1}{h^2} (f(x_0) - 2f(x_1) + f(x_2)).
\]  

(2.78)

It is interesting to note that, like the two-point approximation to the first derivative (Eq. (2.19)), this approximation is the same irrespective of which node you use for the evaluation. In other words, the approximation is the same for each node of the three node series.

Error in approximation

From Eq. (2.70) we have

\[
\frac{dE(x)}{dx} = \frac{1}{(n + 1)!} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \right) \bigg|_{x=\xi} \frac{d}{dx} \left[ \Pi_{i=0}^{n} (x - x_i) \right]
\]  

(2.79)

thus, upon differentiation we have

\[
\frac{d^2 E(x)}{dx^2} = \frac{d}{dx} \left[ \frac{1}{(n + 1)!} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \right) \bigg|_{x=\xi} \frac{d}{dx} \left[ \Pi_{i=0}^{n} (x - x_i) \right] \right]
\]  

(2.80)

or (see page 34)

\[
\frac{d^2 E(x)}{dx^2} = \frac{1}{(n + 2)!} \left( \frac{d^{(n+2)} f(x)}{dx^{(n+2)}} \bigg|_{x=\eta} \right) \frac{d}{dx} \left[ \Pi_{i=0}^{n} (x - x_i) \right]
\]
\[
+ \frac{1}{(n + 1)!} \left( \frac{d^{(n+1)} f(x)}{dx^{(n+1)}} \bigg|_{x=\xi} \right) \frac{d^2}{dx^2} \left[ \Pi_{i=0}^{n} (x - x_i) \right].
\]  

(2.81)
If we select \( n = 2 \) this equation becomes

\[
\frac{d^2 E(x)}{dx^2} = \frac{1}{(2 + 2)!} \left( \frac{d^{2+2} f(x)}{dx^{2+2}} \right) \bigg|_{x = \eta} \frac{d}{dx} \left[ \Pi_{i=0}^2 (x - x_i) \right]
+ \frac{1}{(2 + 1)!} \left( \frac{d^{2+1} f(x)}{dx^{2+1}} \right) \bigg|_{x = \xi} \frac{d^2}{dx^2} \left[ \Pi_{i=0}^2 (x - x_i) \right]
\]

(2.82)

which simplifies to

\[
\frac{d^2 E(x)}{dx^2} = \frac{1}{4!} \left( \frac{d^4 f(x)}{dx^4} \right) \bigg|_{x = \eta} \frac{d}{dx} \left( x - x_0 \right) (x - x_1) (x - x_2)
+ \frac{1}{3!} \left( \frac{d^3 f(x)}{dx^3} \right) \bigg|_{x = \xi} \frac{d^2}{dx^2} \left[ (x - x_0) (x - x_1) (x - x_2) \right].
\]

(2.83)

At the node \( x_1 \) this expression becomes

\[
\left. \frac{d^2 E}{dx^2} \right|_{x_1} = \frac{1}{4!} \left( \frac{d^4 f(x)}{dx^4} \right) \bigg|_{x = \eta} \left( (x_1 - x_0) (x_1 - x_1) + (x_1 - x_1) (x_1 - x_2) + (x_1 - x_0) (x_1 - x_2) \right)
+ \frac{1}{3!} \left( \frac{d^3 f(x)}{dx^3} \right) \bigg|_{x = \xi} \left[ 2 (x_1 - x_0) + 2 (x_1 - x_1) + 2 (x_1 - x_2) \right]
\]

(2.84)

which simplifies to

\[
\left. \frac{d^2 E}{dx^2} \right|_{x_1} = \frac{1}{4!} \left( \frac{d^4 f(x)}{dx^4} \right) \bigg|_{x = \eta} \left[ (x_1 - x_0) (x_1 - x_2) \right]
+ \frac{1}{3!} \left( \frac{d^3 f(x)}{dx^3} \right) \bigg|_{x = \xi} \left[ 2 (x_1 - x_0) + 2 (x_1 - x_2) \right]
\]

(2.85)  

(2.86)

and finally becomes

\[
\left. \frac{d^2 E}{dx^2} \right|_{x_1} = \left\{ \frac{1}{4!} \left( \frac{d^4 f(\eta)}{dx^4} \right) \bigg|_{x = \eta} (-h^2) + \frac{2}{3!} \left( \frac{d^3 f(x)}{dx^3} \bigg|_{x = \xi} \right) (0) \right\} = O(h^2).
\]

(2.87)

**Example of the Solution of a Second-Order Equation**

In this example we will show how to solve a second-order equation using a finite difference approach. In this spirit, consider the transport of a contaminant \( c \) being transported by a fluid with velocity \( v \) and experiencing dispersion with dispersion coefficient \( D \). Let the interval of interest be \( x = [0, 1] \). The governing equation is

\[
D \frac{d^2 c}{dx^2} - v \frac{dc}{dx} = 0.
\]

(2.88)

Assume that the concentration at \( x = 0 \) is 1 and at \( x = 1 \) is 0. Further assume that we are going to use three nodes located at \( x_0 = 0, x_1 = 0.5, \) and \( x_2 = 1.0 \). Because the first and last nodes in the three node sequence are known, we only need to solve
for the center node. This requires one equation in one unknown, namely the unknown value at \( x = 0.5 \), which is located at node 1 in the sequence.

Using centered finite difference approximations for each term in Eq. (2.88), we obtain for the equation at node \( i \)

\[
D \frac{c_{i+1} - 2c_i + c_{i-1}}{h^2} - v \frac{c_{i+1} - c_{i-1}}{2h} = 0.
\]  

(2.89)

For the special case of \( n = 1 \), this expression becomes

\[
D \frac{c_2 - 2c_1 + c_0}{h^2} - v \frac{c_2 - c_0}{2h} = 0.
\]  

(2.90)

We can write the matrix equation as

\[
\begin{bmatrix}
\frac{D}{h^2} & \frac{v}{2h} & 0 & 0 & 0 \\
0 & \frac{D}{h^2} & \frac{-2D}{h^2} & 0 & 0 \\
0 & 0 & \frac{D}{h^2} & \frac{v}{2h} & 0 \\
0 & 0 & 0 & \frac{D}{h^2} & \frac{v}{2h} \\
\end{bmatrix}
\begin{bmatrix}
1 \\
c_0 \\
c_1 \\
c_2 \\
0
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}.
\]  

(2.91)

where the first and last rows exist only to incorporate the boundary conditions. More specifically, if you multiply the column of \( c_i \) values by the first row of the matrix you obtain \( c_0 = 1 \), which is the boundary condition at \( x = 0 \). A similar argument can be made for the representing the boundary condition at \( x = 1 \) by using the last row of the matrix equation.

Now assume for convenience that \( D = v = 1 \) and substitute into Eq. (2.91) which now becomes

\[
\begin{bmatrix}
1 & \frac{1}{1/4 + 1} & \frac{0}{1/4} & \frac{0}{1/4 - 1} & \frac{0}{1} \\
0 & \frac{1/4}{1/4} & \frac{-2D}{1/4} & \frac{0}{1/4 - 1} & \frac{0}{1} \\
0 & \frac{0}{1/4} & \frac{D}{1/4 - 1} & \frac{0}{1} & \frac{0}{1} \\
0 & \frac{0}{1/4 - 1} & \frac{0}{1} & \frac{D}{1/4} & \frac{v}{2h} \\
\end{bmatrix}
\begin{bmatrix}
c_0 \\
c_1 \\
c_2 \\
0 \\
0
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}.
\]  

(2.92)

Focusing now on the equation defined by the second row of Eq. (2.92) we obtain through matrix multiplication

\[
\begin{bmatrix}
1 \\
\frac{0}{1/4} \\
\frac{0}{1/4 - 1} \\
\frac{0}{1} \\
\frac{0}{1/4} \\
\end{bmatrix}
\begin{bmatrix}
c_0 \\
c_1 \\
c_2 \\
0 \\
0
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}.
\]  

(2.93)

which can now be solved for the value of \( c \) at \( x = 0.5 \), that is \( c_1 \) and the result is

\[
c_1 = \left( \frac{1}{8} \right) \left( \frac{5}{1} \right) \left( 1 \right) + \left( \frac{3}{8} \right) \left( 0 \right) = \frac{5}{8}.
\]  

(2.94)

Now, just out of curiosity, let us see what happens as we change the coefficient \( D \). Writing the solution for \( c_1 \) symbolically using Eq. (2.91) we have (noting that \( c_2 = 0 \) via the boundary condition at \( x = 1 \))

\[
c_1 = \left( \frac{1/4}{2D} \right) \left( \frac{D}{1/4} + v \right) c_0
\]  

(2.95)

or

\[
c_1 = \frac{1}{2} + \frac{v}{8D}.
\]  

(2.96)

Notice that as \( D \to \infty \), \( c_1 \to 0.5 \) and as \( D \to 0 \), \( c_1 \to \infty \). The reason for \( c_1 \to \infty \) is that the right hand boundary condition is inappropriate and causes a conflict. The solution approaches 1 everywhere when \( D = 0 \) but this is in conflict with \( c_2 = 0 \) which
is the right-hand boundary condition. As we noted in the previous example, having two boundary conditions specified for a first-order equation (remember \(D = 0\) is assumed) is incorrect. The solution for \(c_1\) reflects that fact. In practice Eq. (2.88) is difficult to solve when the ratio of \(v/D\) becomes large for a variety of reasons as we will see later.

### 2.5 Chapter Summary

The polynomials introduced in the preceding chapter are used to create discrete approximations to continuous derivatives. The error in these approximations is introduced and discussed. The resulting discrete approximations to first and second order derivatives are used to represent differential equations and the result is a set of algebraic equations which, when solved, generate an approximate solution to the differential equation. By using different strategies for approximating the continuous derivatives, the discrete equations can have different numbers of nodal locations in their approximations and concomitant differences in the order of their errors. As an illustration of the use of the finite difference methods that arise using this strategy, an example differential equation is solved employing this approach.

### 2.6 Problems

1. Consider the equation

\[
\frac{d^2 u}{dx^2} + u = 0 \quad x \in [0, 1] \tag{2.97}
\]

\[
u(0) = 1 \tag{2.98}
\]

\[
\left. \frac{du}{dx} \right|_{x=1} = 0. \tag{2.99}
\]

Using three nodes located at \(x = 0\), \(x = 0.5\), and \(x = 1.0\), find the value of \(u(0.5)\). Note that there is more than one option for the representation of the variable \(u_1\) located at \(x = 0.5\).

2. Consider the following equation and associated boundary condition

\[
\frac{dv}{dx} + v = 0 \tag{2.100}
\]

\[
v(0) = 3. \tag{2.101}
\]

The analytic solution is

\[
v = 3e^{-x} \tag{2.102}
\]

so

\[
v(1) = \frac{3}{e^1} = \frac{3}{2.718} = 1.1 \tag{2.103}
\]

\[
v(2) = \frac{3}{e^2} = \frac{3}{7.39} = 0.405. \tag{2.104}
\]

Given the mesh shown in Fig. 2.2, calculate \(v(1)\) using a backward difference scheme and compare your results to the exact values.
2.6. PROBLEMS

3. Given Eqs. (2.105)-(2.107) and a mesh made up of three nodes located at $x_0 = 0$, $x_1 = 0.5$, and $x_2 = 1.0$, use a second order correct finite difference scheme to determine the value of $c$ at $x = 0.5$. Assume $D = 2$ and $v = 1$:

$$D \frac{d^2 u}{dx^2} + u = 0 \quad x \in [0, 1]$$ (2.105)

$$u(0) = 1$$ (2.106)

$$u(1) = 1.$$ (2.107)

4. Given the error expression

$$\frac{dE}{dx}\bigg|_{x_j} = \frac{1}{(n+1)!} \left( \frac{d^{(n+1)} f(\xi)}{dx^{(n+1)}} \right) \frac{d}{dx} \left[ \prod_{i=0}^{n} (x - x_i) \right]_{x_j}$$ (2.108)

determine the error bounds for $\frac{d}{dx} \left( \sin(\theta) \right)$ at $\pi/2$ and $\pi$, using the mesh shown in Fig. 2.3 and an approximation using linear Lagrange polynomials ($n = 1$). Are they equivalent and should they be?

5. Consider the physical system illustrated in Fig. 2.4. Fluid is entering the pipe at a rate $Q = x$ per unit length. The velocity entering the pipe is $v = 3$. What
is the velocity of the pipe along its length? The governing equation, boundary condition, and flux are given as:

$$\frac{dv}{dx} = Q$$  \hspace{1cm} (2.109)

$$v(0) = 3$$  \hspace{1cm} (2.110)

$$Q = x.$$  \hspace{1cm} (2.111)

Compare your solution to the analytical one, namely

$$v = \frac{x^2}{2} - 3.$$  \hspace{1cm} (2.112)

Figure 2.4: Problem formulation where fluid enters a pipe at $x_0$ and there is a source $Q = x$ along the length of the pipe. The input velocity is $v = 3$.

6. In Fig. 2.5 is displayed the water-level response to a pumping well drawing from a groundwater reservoir of thickness $l$. The well produces a drop in water level, $h$. The water level drops most at the well and less away from the well. It is assumed that at some distance $r_e$ the water level is not disturbed and remains at height $H_e$. The equation describing the drop in water level is given by

$$Q = \frac{K dH}{dr} 2\pi rl$$  \hspace{1cm} (2.113)

where $Q$ is the well discharge. The solution to this equation is

$$H_e - h = \frac{Q}{2\pi Kl} \ln \frac{r_e}{r}.$$  \hspace{1cm} (2.114)

Assume that the thickness $l$ is 10, that $r_e = 100$, $H_e = 10$, $Q = 1$, and $K = 0.1$. Calculate the value of $h$ at $r = 50$ using a finite difference scheme based on linear Lagrange polynomials and a sequence of element sizes beginning with 50 (3 nodes), then 25 (5 nodes), then 12.5 (9 nodes) and finally 6.25 (17 nodes). Compute the magnitude of the error $h - \hat{h}$, where $\hat{h}$ is the approximate solution at $r = 50$. Plot the logarithm of the error at that point for each scheme on the vertical axis and the logarithm of the element size on the horizontal axis. Calculate the slope of this line. This is an estimate of the order of the global error. Given the method that you have used to approximate the derivative, how does the local error compare to the global error?
7. The convection diffusion equation is one of the most often encountered in science and engineering. Solve the following problem that uses Eq. (2.115). The problem is defined for the interval $X \in [0, 1]$ by

$\frac{-d^2\phi}{dX^2} + P_e \frac{d\phi}{dX} = 0 \quad X \in [0, 1] \tag{2.115}$

$\phi(0) = 1 \tag{2.116}$

$\phi(1) = 0 \tag{2.117}$

where $X = \frac{x}{L}$, $\phi$ is a measure of concentration and the Peclet number $P_e = \frac{\rho u L}{\Gamma}$, $\rho$ is fluid density, $u$ is fluid velocity, $\Gamma$ is the diffusivity, and $L$ is a characteristic length. The goal of this exercise is to see under what values of $P_e$ the numerical solution of this equation becomes problematic. Notice that by increasing $P_e$ you are effectively increasing the importance (influence) of the convective term at the expense of the diffusive term.

1. The analytical solution to this problem is

$\phi = \phi(0) + (\phi(1) - \phi(0)) \frac{\exp(P_e X) - 1}{\exp(P_e) - 1}. \tag{2.118}$

Proceed as follows:

(a) Replace the value of $P_e$ with the numerical equivalent $P_{eh} = \frac{\rho u h}{\Gamma}$ that uses the spacing $h$ as the value of $L$. Notice that the $h$ in the $P_{eh}$ term is obtained by multiplying your difference equation through by $h$, so the $h$ does not
(b) Use two different numerical approximations.
   i. the first is a central difference approximation for both the first and second derivatives.
   ii. the second is a central difference approximation for the second derivative and a backward difference for the first derivative.

(c) Use two different values of \( P_e \):
   i. one case is with \( P_h = 1 \).
   ii. one case is with \( P_h = 3 \).

Since you are using only one value of \( h \) in all your simulations, by changing \( P_h \), you are really changing either \( u \) or \( \Gamma \).

d. Using a mesh with 17 nodes, calculate the value of \( \phi_i \) at each node with the boundary conditions indicated. Do this for the two values of \( P_h \) and each numerical approximation.

e. Plot these four solutions along with their analytical equivalent (using \( P_e = P_h \)). Theory suggests that with a value of \( P_e \) greater than 2, you should have problems for the case where the central difference approximation is used for the first derivative and all the other cases should be fine. Also, theoretically, the solution for \( P_h = 1 \) should be more accurate when the central difference approximation for the first derivative is used (why should this be true?).

8. Given the \( \sin(\theta) \) function shown in Fig. 2.6 and linear polynomials determine the value of the error range of \( \frac{dE(\theta)}{d\theta} \) for the derivative of \( \sin(\theta) \) at \( \theta = 0 \). Compare this error to the actual error you get when you attempt to approximate \( \frac{d\sin(\theta)}{d\theta} \bigg|_{\theta=0} \). Is the actual error within the range determined from \( \frac{dE(\theta)}{d\theta} \)?

![Figure 2.6: Sin function from 0 \( \leq \theta \leq \pi \).](image)

9. Use quadratic (fourth degree) Lagrange polynomials to approximate \( \sin \theta \) using nodes at \( 0, \pi/2, \pi, (3/2)\pi \) and \( \pi \) (see Fig. 2.7). Determine the value of \( \sin \theta \) at \( (5/4)\pi \). Compare the estimate to the value obtained using the quadratic Lagrange polynomial. Is it better; why?
2.6. PROBLEMS

Figure 2.7: Sin ($\theta$) curve with measured points indicated by black dots and interpolated values indicated by grey dots.

10. Given the equation

$$\frac{\partial u}{\partial x} + \frac{\partial^2 u}{\partial x^2} = 0$$

(2.119)

determine whether the expression

$$\frac{u_{i+1} - u_i}{h} + \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} = 0$$

(2.120)

is consistent.

11. Show why you cannot use linear Lagrange polynomial functions and two nodes to approximate a term of the form $\frac{\partial^2 u}{\partial x^2}$.

12. Reconsider the problem presented earlier in this chapter, however with a slight modification. Again we let the velocity entering the pipe be $v = 3$. The governing equation, boundary condition, and known flux $Q$ are given as earlier, but $A$ is now defined as a function of $x$ (see Fig. 2.8)., The problem now is:

$$A \frac{dv}{dx} = Q \quad x \in [0, 3]$$

(2.121)

$$v(0) = 3,$$

(2.122)

$$Q = 1,$$

(2.123)

$$A = 1 - x/6.$$  

(2.124)

The analytical solution is

$$v = -6 \ln (1 - x/6) + 3.$$  

(2.125)
Figure 2.8: Physical system for problem with linearly decreasing tube diameter.

The wrinkle here is how to handle the coefficient $A$. One approach is to evaluate the coefficient at each node, that is, $A = 1 - x_i/6$.

Write a computer code to calculate the solution at each of the nodes and evaluate the error at each node. Plot the error as a function of the distance along $x$. Repeat the calculation and analysis using seven equally spaced nodes. Plot the error at each of the nodes for which you have values from each calculation. As a measure of error, take the largest error among all the nodes for a particular discretization as the measure of error for that discretization.
Bibliography


Chapter 3

Numerical Integration

As we move into a discussion of numerical methods that require the evaluation of integrals, we discover that there is a need to understand some of the elementary aspects of numerical integration. There is a vast literature on this topic and this chapter addresses only those aspects we need later to assist in understanding the finite element and finite volume methods.

3.1 Newton-Cotes Quadrature Formula

In this subsection we use the techniques we have developed to date to examine the family of integration methods known as Newton-Cotes Quadrature formulae. We begin with reviewing some of the material presented earlier.

3.1.1 Lagrange Interpolation

From the general formula for approximating a function using Lagrange polynomials we have

\[ f(x) = \sum_{j=0}^{n} \ell_j(x) f(x_j) + \frac{1}{(n+1)!} \Pi_{i=0}^{n} (x - x_i) \frac{d^{n+1}f(\xi)}{dx^{n+1}} \quad \xi \in [a, b]. \]  

(3.1)

Let us now integrate Eq. (3.1) over an interval from \( a \) to \( b \). For \( n \) even one can show that (Ralston, 1965) [9]

\[
\int_{a}^{b} f(x) \, dx = \sum_{j=0}^{n} \int_{a}^{b} \ell_j^n(x) f(x_j) \, dx \\
+ \frac{1}{(n+2)!} \frac{d^{n+2}f(\eta)}{dx^{n+2}} \int_{a}^{b} x\Pi_{i=0}^{n} (x - x_i) \, dx
\]

(3.2)

and for \( n \) odd
\[ \int_a^b f(x) \, dx = \sum_{j=0}^{n} \int_a^b \ell_j^n(x) f(x_j) \, dx + \frac{1}{(n+1)!} \frac{d^{n+1} f(\eta)}{dx^{n+1}} \int_a^b \Pi_i^n(x-x_i) \, dx. \]  

(3.3)

We will now see that from these two expressions, we can formulate the commonly encountered quadrature formulae for numerical integration.

### 3.1.2 Trapezoidal Rule

Let us start with the simplest quadrature formula, namely the case when \( n = 1 \). We consider the interval \( x \in [0, h] \) with, as earlier, \( h = x_i - x_{i-1} \), \( x = ih \) and \( x_i \) are nodal locations along the \( x \) axis. Since \( n = 1 \) is odd, we use Eq. (3.3), that is

\[ \int_0^h f(x) \, dx = \int_0^h \left[ \ell_0(x) f(x_0) + \ell_1(x) f(x_1) \right] \, dx + \frac{1}{(2)!} \frac{d^2 f(\eta)}{dx^2} \int_0^h (x-x_0) (x-x_1) \, dx. \]  

(3.4)

The next step is to substitute our definition of the linear Lagrange polynomial and recognize that \( f(x_i) \) is a number (the value of \( f(x) \) at the specific locations \( x_i \)) and can be taken out of the integral, we obtain

\[ \int_0^h f(x) \, dx = \left( \int_0^h \frac{h-x}{h} \, dx \right) f(x_0) + \left( \int_0^h \frac{x}{h} \, dx \right) f(x_1) + \frac{1}{(2)!} \frac{d^2 f(\eta)}{dx^2} \int_0^h (x-x_0) (x-h) \, dx. \]  

(3.5)

Now we perform the integrations appearing on the right-hand side of Eq. (3.5) and obtain

\[ \left( \int_0^h \frac{h-x}{h} \, dx \right) = \left( x - \frac{x^2}{2h} \right) \bigg|_0^h = \left( h - \frac{h^2}{2h} \right) = \frac{h}{2} \]  

(3.6)

\[ \left( \int_0^h \frac{x}{h} \, dx \right) = \left( \frac{x^2}{2h} \right) \bigg|_0^h = \frac{h}{2} \]  

(3.7)

\[ \frac{1}{(2)!} \frac{d^2 f(\eta)}{dx^2} \int_0^h (x-x_0) (x-h) \, dx \]

\[ = \frac{1}{(2)!} \frac{d^2 f(\eta)}{dx^2} \int_0^h (x^2-hx) \, dx \]

\[ = \frac{1}{(2)!} \frac{d^2 f(\eta)}{dx^2} \left( \frac{x^3}{3} - \frac{hx^2}{2} \right) \bigg|_0^h \]

\[ = O \left( h^3 \right). \]  

(3.8)

Substitution of these integrals into 3.5 provides us with the final form for the integral \( \int_0^h f(x) \, dx \).
3.1. NEWTON-COTES QUADRATURE FORMULA

\[ \int_0^h f(x) \, dx = \frac{h}{2} f(0) + \frac{h}{2} f(h) + O(h^3). \]  

(3.9)

Equation 3.9 tells us that we can obtain an approximation to the integral \( \int_0^h f(x) \, dx \) as the weighted sum of the values at the end points of the interval, namely \( f(0) \) and \( f(h) \) with weighting coefficients \( \frac{h}{2} \). An alternative interpretation is that the integral is approximated by the arithmetic sum of the values at the ends of the interval, multiplied by the interval value \( h \). Now let us extend the concept to the case of three nodes.

3.1.3 Simpson’s Rule

In the case of three nodes, \( n = 2 \). If we assume \( h \) to be a constant, the interval over which we are doing our integration becomes \( x \in [0, 2h] \). Since \( n \) is even, we use Eq (3.2), that is

\[
\begin{align*}
\int_0^{2h} f(x) \, dx &= \int_0^{2h} \left[ \ell_0^2(x) \, f(x_0) + \ell_1^2(x) \, f(x_1) + \ell_2^2(x) \, f(x_2) \right] \, dx \\
&+ \frac{1}{(2+2)!} \int_0^{2h} \frac{d^4 f(\eta)}{dx^4} \int_0^{2h} x (x-x_0)(x-x_1)(x-x_2) \, dx. \\
&= 1/2h^2 \int_0^{2h} \left[ \frac{x^3}{3} - \frac{x_1^2}{2} - \frac{x_2^2}{2} + x_1 x_2 \right] \, dx \int_0^{2h} \left[ \frac{x^3}{3} - \frac{x_1^2}{2} - \frac{x_2^2}{2} + x_1 x_2 \right] \, dx \int_0^{2h} \left[ \frac{x^3}{3} - \frac{x_1^2}{2} - \frac{x_2^2}{2} + x_1 x_2 \right] \, dx.
\end{align*}
\]

(3.10)

To get a flavor for the process involved in evaluating the terms in Eq. (3.10), consider the first term in the first integrand appearing on the right-hand side of this equation, viz.

\[
\begin{align*}
\int_0^{2h} \left[ \ell_0^2(x) \, f(x_0) \right] \, dx &= \left( \int_0^{2h} \left[ \ell_0^2(x) \right] \, dx \right) \, f(x_0) \\
&= \left( \int_0^{2h} \left[ \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} \right] \, dx \right) \, f(x_0) \\
&= \left( \int_0^{2h} \left[ \frac{x^2-x_1 x-x_2 x+x_1 x_2}{(-h)(-2h)} \right] \, dx \right) \, f(x_0) \\
&= \left( \frac{1}{2h^2} \left[ \frac{x^3}{3} - \frac{x_1^2}{2} - \frac{x_2^2}{2} + x_1 x_2 \right] \right) \, f(x_0) \\
&= \left( \frac{1}{2h^2} \left[ \frac{16h^3}{6} - \frac{12h^3}{6} - \frac{24h^3}{6} + \frac{24h^3}{6} \right] \right) \, f(x_0) \\
&= \frac{1}{2h^2} = \frac{2}{3} f(x_0).
\end{align*}
\]

(3.11)
The integral in the error term becomes

\[
\int_0^{2h} x^2 (x - h)(x - 2h) \, dx = \int_0^{2h} x^2 (x^2 - 3hx + 2h^2) \, dx = \int_0^{2h} (x^4 - 3hx^3 + 2x^2h^2) \, dx = \left( \frac{x^5}{5} - \frac{3}{4}hx^4 + \frac{2h^2}{3}x^3 \right)_0^{2h} = \frac{32h^5}{5} - \frac{48h^5}{4} + \frac{16h^5}{3} = -\frac{4}{15}h^5.
\]  

(3.12)

Combination of these integral terms yields

\[
\int_0^{2h} f(x) \, dx = \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)] + \frac{1}{4!} \left( \frac{-4}{15} \right) h^5 \frac{d^4 f(\eta)}{dx^4} \quad \text{for} \quad \frac{1}{4!} \left( \frac{-4}{15} \right) h^5 \frac{d^4 f(\eta)}{dx^4} 
\]

\[
= \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)] + O(h^5).
\]  

(3.13)

Equation 3.13 states that we can obtain a value for the integral \( \int_0^{2h} f(x) \, dx \) using three nodal values of \( f(x) \), that is \( f(x_0) \), \( f(x_1) \), and \( f(x_2) \) that is \( O(h^5) \). Thus we observe that the use of additional nodes increases the accuracy of the numerical approximation; of course it also requires more computational effort.

### 3.1.4 General Form

The general form of the above is

\[
\int_{x_0}^{x_n} f(x) \, dx = \sum_{j=0}^{n} H_j f(x_j) + E
\]

\[
H_j = hAW_j
\]

\[
E = E_1 h^{(k+1)} \frac{d^k f(\eta)}{dx^k}
\]  

(3.14)

where

\[
k = n + 1 \quad \text{if} \ n \ \text{is odd}
\]

\[
k = n + 2 \quad \text{if} \ n \ \text{is even.}
\]  

(3.15)

(3.16)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( A )</th>
<th>( W_0 )</th>
<th>( W_1 )</th>
<th>( W_2 )</th>
<th>( W_3 )</th>
<th>( E_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/2</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>-1/12</td>
</tr>
<tr>
<td>2</td>
<td>1/3</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td></td>
<td>-1/90</td>
</tr>
<tr>
<td>3</td>
<td>3/8</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>-3/80</td>
</tr>
</tbody>
</table>
3.1.5 Example using Simpson’s Rule

As an example, let us perform the integration as seen in Fig. 3.1. The exact value for this integral is

$$\int_1^3 (x^2 + x + 1) \, dx = 14.666. \quad (3.18)$$

Let us use a second degree polynomial such that $n = 2$. Then from Eq. (3.16) we have $k = 2 + 2$. We see from Eq. (3.14) that if we use a second degree polynomial, there should be no error since the fourth derivative of our integrand is zero; let us see if this holds for our example. From Simpson’s rule

$$\int_1^3 (x^2 + x + 1) \, dx$$

$$= h \left[ \frac{1}{3} f(x_0) + \frac{4}{3} f(x_1) + \frac{1}{3} f(x_2) \right]$$

$$= 1 \left[ \left( \frac{1}{3} \right) 3 + \left( \frac{4}{3} \right) 7 + \left( \frac{1}{3} \right) 13 \right]$$

$$= 14.666. \quad (3.19)$$

Thus, as hypothesized, the integration is exact.

Figure 3.1: Function to be integrated and integral value.

3.1.6 Gauss Legendre Quadrature

In this subsection we look at a quite different method of integration that we will need later in our discussion of isoparametric finite element (see Section 8.3 on page 211) and finite volume methods (see Section 9.2.5 on page 256). This method, called Gauss Legendre Quadrature, can be presented at various levels of sophistication. The following approach is consistent with our needs.

Assume the integral of interest can be approximated as

$$\int_a^b f(x) \, dx \approx \sum_{i=1}^n c_i f(x_i), \quad (3.20)$$
where both the coefficients $c_i$ and the locations $x_i$ are unknown. We will now determine, using two terms of this series, the values of $c_1$, $c_2$, $x_1$, and $x_2$ that exactly integrates an equation of the form

$$f(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3.$$ \hspace{1cm} (3.21)

Substitution of Eq. (3.21) into Eq. (3.20) gives us

$$\int_a^b f(x) \, dx = \int_a^b (a_0 + a_1 x + a_2 x^2 + a_3 x^3) \, dx.$$ \hspace{1cm} (3.22)

The exact solution to this equation is

$$\int_a^b f(x) \, dx = a_0 (b-a) + a_1 \left( \frac{b^2 - a^2}{2} \right) + a_2 \left( \frac{b^3 - a^3}{3} \right) + a_3 \left( \frac{b^4 - a^4}{4} \right).$$ \hspace{1cm} (3.23)

From Eq. (3.20) we have, for $n = 2$

$$\int_a^b f(x) \, dx \approx \sum_{i=1}^{2} c_i f(x_i) = c_1 f(x_1) + c_2 f(x_2).$$ \hspace{1cm} (3.24)

Using now the definition of $f(x_i)$ from Eq. (3.21), we obtain for the two values of $x_i$

$$c_1 f(x_1) + c_2 f(x_2) = c_1 \left( a_0 + a_1 x_1 + a_2 x_1^2 + a_3 x_1^3 \right) + c_2 \left( a_0 + a_1 x_2 + a_2 x_2^2 + a_3 x_2^3 \right).$$

Rearrangement of the terms in this expression yields

$$c_1 f(x_1) + c_2 f(x_2) = a_0 \left( c_1 + c_2 \right) + a_1 \left( c_1 x_1 + c_2 x_2 \right) + a_2 \left( c_1 x_1^2 + c_2 x_2^2 \right) + a_3 \left( c_1 x_1^3 + c_2 x_2^3 \right).$$

We now have two equations that both must be satisfied, these are Eq. (3.23) and Eq. (3.24). For this to be true

$$c_1 + c_2 = b - a$$
$$c_1 x_1 + c_2 x_2 = \frac{b^2 - a^2}{2}$$
$$c_1 x_1^2 + c_2 x_2^2 = \frac{b^3 - a^3}{3}$$
$$c_1 x_1^3 + c_2 x_2^3 = \frac{b^4 - a^4}{4}.$$ \hspace{1cm} (3.25)

Equation (3.25) represent four non-linear equations in four unknowns. Solution of these equations yields

$$c_1 = \frac{b - a}{2}$$
$$c_2 = \frac{b - a}{2}$$
$$x_1 = \frac{b - a}{2} \left( -\frac{1}{\sqrt{3}} \right) + \frac{b + a}{2}$$
$$x_2 = \frac{b - a}{2} \left( \frac{1}{\sqrt{3}} \right) + \frac{b + a}{2}.$$ \hspace{1cm} (3.26)
Substitution of these values into Eq. (3.24) yields the integration formula we seek.

In application, the general form of the approximation is modified to accommodate integration limits defined by \(a = -1, b = +1\). This is the form we will use later in the book. In this case, the coefficients in Eq. (3.26) become

\[
\begin{align*}
    c_1 &= \frac{b - a}{2} = 1 \\
    c_2 &= \frac{b - a}{2} = 1 \\
    x_1 &= -\frac{1}{\sqrt{3}} \\
    x_2 &= \frac{1}{\sqrt{3}}.
\end{align*}
\] (3.27)

In Table 3.29, are tabulated the values of \(c_i\) and \(x_i\) for \(n = 1, 2, 3, 4\). In our work we will need to calculate multidimensional integrals. The formula for two dimensions is a straightforward extension of the one-dimensional case and is for \(\int_{-1}^{1} \int_{-1}^{1} f(x, y) dx dy\)

\[
\int_{-1}^{1} \int_{-1}^{1} f(x, y) dx dy = \sum_{i=1}^{n} \sum_{j=1}^{m} c_i d_j f(x_i, y_j) (3.28)
\]

where \(n\) and \(m\) are the number of integration points in \(x\) and \(y\), \(c_i\) and \(d_j\) are weighting coefficients and \(f(x_i, y_j)\) is the value of the integrand at the specific points \(x_j \in [-1, 1], y_i \in [-1, 1]\). In this formula, we introduce \(d_j\) and \(y_j\) to denote values equivalent to \(c_j\) and \(x_j\) but taken in the \(y\) coordinate direction. In our above example, we considered a cubic polynomial which we observed was exactly integrated with an \(n = 2\). This is consistent with information presented in Table 3.29. In general, the polynomial that will be exactly integrated by \(n\) points is of degree of \(2n - 1\). Notice that there are + and − values for \(x_i\) so the number of points being considered for, say, \(n = 2\), is two and for \(n = 3\) is three.

<table>
<thead>
<tr>
<th>(i)</th>
<th>Polynomial degree</th>
<th>(\pm x_i)</th>
<th>(c_i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Linear</td>
<td>0.000</td>
<td>2.000</td>
</tr>
<tr>
<td>2</td>
<td>Cubic</td>
<td>0.577</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>Quintic</td>
<td>0.775</td>
<td>0.556</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000</td>
<td>0.889</td>
</tr>
<tr>
<td>4</td>
<td>Septimal</td>
<td>0.861</td>
<td>0.348</td>
</tr>
</tbody>
</table>
|      |                  | 0.340      | 0.652  | (3.29)

**Example Gauss Integration**

Consider the integral we used as an example in Subsubsection 3.1.5, that is

\[
\int_{-1}^{1} (z^2 + z + 1) \, dx = 14.666. \quad (3.30)
\]
**Step One: Change Variable Limits**, Change the limits of integration to \([-1, 1]\) using the following equation:

\[ x = \frac{1}{2} (2z - 4) \quad (3.31) \]

and verify the transformation is correct.

If \(x = 1\),

\[ x = \frac{1}{2} (2 - 4) = -1. \quad (3.32) \]

If \(x = 3\),

\[ x = \frac{1}{2} (6 - 4) = 1. \quad (3.33) \]

Now, the integral becomes

\[ \int_{-1}^{1} (z^2 + z + 1) \, dz = \int_{-1}^{1} (x^2 + 5x + 7) \, dx. \quad (3.34) \]

This is the interval from \(-1\) to \(1\), we stated that we needed to apply Gauss Legendre Quadrature as it is used in this book. For convenience define

\[ X = (x^2 + 5x + 7). \quad (3.35) \]

**Step Two: Evaluate Series**

\[ \gamma = \sum_{j=1}^{n} c_j X(x_j). \quad (3.36) \]

If we choose \(n = 2\), we have

\[ \gamma = \sum_{j=1}^{2} c_j X(x_j) \]
\[ = 1 \times \left[ (0.577)^2 + 5 \times 0.577 + 7 \right] + 1 \left[ (0.577)^2 - 5 \times 0.577 + 7 \right] \]
\[ = 14.666. \quad (3.37) \]

The integration is exact because \(n = 2\) will exactly integrate a cubic and our function is quadratic.

### 3.2 Chapter Summary

Polynomial approximation theory is used in this brief chapter to integrate functions numerically. The two methods discussed are the Newton-Cotes quadrature and the Gauss-Legendre quadrature. Common formulae, such as the trapezoidal rule and Simpson’s rule are described and examples of their application provided. An example of the application of the Gauss-Legendre method is also provided. The material in this chapter demonstrates that integration of polynomials can be done easily using numerical strategies based upon polynomial approximation theory.
3.3 Problems

1. Use the trapezoidal rule to evaluate the integral of $\sin \theta$ from 0 to $\pi$ and check to see if the error obtained is consistent with the theoretical limits.

2. Use Simpson’s rule to integrate $\sin \theta$ from 0 to $\pi$, compute the error in the numerical integration and compare this error to the theoretical limits.

3. Develop the one point integration formula for the linear equation $f(x) = a_0 + a_1x$.

4. What is the value of the truncation error generated when one applies Simpson’s rule

$$
\int_0^{2h} f(x) \, dx = h \left( \frac{1}{3} f(x_0) + \frac{4}{3} f(x_1) + \frac{1}{3} f(x_2) \right) - \frac{1}{90} h^5 \frac{d^4 f(\eta)}{dx^4} \quad (3.38)
$$

when $f(x)$ is given by $f(x) = a + bx + cx^2 + dx^3$?

5. Use Simpson’s rule to evaluate the integral in Eq. (3.39) and check to see if the error obtained is consistent with the theoretical limits.

$$
\int_1^3 (x^2 + x + 1) \, dx = 14.666 
$$

Simpson’s rule is given by Eq. (3.38).
Bibliography


Chapter 4

Initial Value Problems

In this section we are interested in expanding the idea of solving differential equations using the finite difference method we first introduced in Chapter 2 on page 2. More specifically we will focus in this chapter on solving ordinary differential equations such as we first encountered in a preliminarily way in Subsection 2.4.2. We will begin this section with equations of the general form

\[
\frac{du}{dt} = f(u, t) \quad t \in [a, b].
\] (4.1)

Since Eq. (4.1) is a first-order equation, it requires one auxiliary condition for its solution, in this case an initial condition since the independent variable is time. A suitable condition is of the form

\[
u(t_0) = u_0,
\] (4.2)

where \(t_0\) is the initial time and \(u_0\) is a specified known value of \(u(t)\). It is assumed that \(f(u, t)\) is also defined and continuous for all \(t \in [a, b]\). Formulating a forward difference approximation of \(\frac{du}{dt}\) appearing in Eq. (4.1) yields (see Section 2.3 on page 37)

\[
\left. \frac{du}{dt} \right|_{t_i} = \frac{u(t_{i+1}) - u(t_i)}{h} + \frac{1}{2!} \frac{d^2u(x)}{dt^2} \bigg|_{x=\xi} (t_i - t_{i+1})
\] (4.3)

where

\[
h = t_{i+1} - t_i
\] (4.4)

Note that we know this is a forward difference because our point of reference for the approximation is at \(t_i\), and we are looking forward to the value at \(t_{i+1}\) to create the approximation.

Substitution of Eq. (4.3) into Eq. (4.1) gives

\[
\left. \frac{du}{dt} - f(u, t) \right|_{t_i} = \frac{u(t_{i+1}) - u(t_i)}{h} - f(u, t_i) - \frac{1}{2!} \frac{d^2u(x)}{dt^2} \bigg|_{x=\xi} h
\]

\[
= 0,
\] (4.5)

or, rearranging

\[
u(t_{i+1}) = \nu(t_i) + hf(u(t_i), t_i) + O(h^2).
\] (4.6)
If we neglect the truncation error, we obtain the Euler Forward Integration Method, that is

\[ u_{i+1} = u_i + hf(t_i, u_i) \]  \hspace{1cm} (4.7)

where \( u_i \) is the approximate solution to \( u(t_i) \). Examination of Eq. (4.7) reveals that one can step through time explicitly, one step at a time, that is the value of \( u_{i+1} \) can be obtained from the known values at \( u_i \). The process gets started using \( u_0 \), which is known from the initial condition.

### 4.1 Euler Forward Integration Method Example

Let us now see how we can apply the above concepts. Consider the equation

\[
\frac{du}{dt} = -\frac{0.9}{1 + 2t} u = f(u, t),
\]

\[ u(0) = u_0 = 1, \]

which has the solution ([4], p. 115).

\[ u(t) = (1 + 2t)^{-0.45}. \]  \hspace{1cm} (4.9)

From Euler’s method Eq. (4.7) we obtain

\[ u_{i+1} = u_i + hf(u_i, t_i). \]  \hspace{1cm} (4.10)

Substitution of our definition for \( f(u, t) \) from Eq. (4.8) gives

\[ u_{i+1} = u_i - h \left( \frac{0.9}{1 + 2t_i} \right) u_i. \]  \hspace{1cm} (4.11)

For our example consider

\[ h = 0.02 \]  \hspace{1cm} (4.12)

such that Eq. (4.11) becomes

\[ u_1 = u_0 - 0.02 \left( \frac{0.9}{1 + 2t_0} \right) u_0. \]  \hspace{1cm} (4.13)

After imposing the initial condition, that is

\[ u_0 = 1, \]  \hspace{1cm} (4.14)

we can write Eq. (4.13) for the first time step as

\[ u_1 = 1 - 0.02 \left( \frac{0.9}{1 + (2)(0)} \right) 1 = 0.982. \]  \hspace{1cm} (4.15)

If we continue this process over several time steps with each newly calculated value depending only on the preceding time step, we obtain the values in the following table. Here, we present, column-wise from left to right, the elapsed time at the beginning of each time step, the computed value of \( u_i \), the exact value of \( u(t) \) at time \( t_i \), and the error as calculated as the difference \( u_i - u(t_i) \). Note that the error reported in the fourth column is cumulative and getting larger as time increases.
4.2 Convergence

There are three key properties of numerical approximations like that considered above that need to be addressed. If these three conditions are not met, the numerical approximation is of little practical interest. These properties are convergence, consistency and stability and we will consider each in turn below.

Convergence Definition

A numerical method applied to a given differential equation is termed convergent if, assuming no computer round-off error, the numerical solution approaches the exact solution to the differential equation as $h \to 0$.

$$\lim_{h \to 0} u_i = u(t_i) \quad \text{where} \quad (h) = \frac{t_i - t_0}{i} \quad \forall \ t \in [a, b].$$  \hspace{1cm} (4.16)

The above means that as the time step $h$ gets smaller and smaller, the calculated value $u_i$ approaches the exact solution $u(t_i)$.

With this definition in mind, we now examine the convergence properties of the Euler method for this example. From the governing equation we have

$$\frac{du}{dt} |_{t_i} = f(u(t_i), t_i)$$  \hspace{1cm} (4.17)

and from Taylor’s theorem we know

$$u(t_{i+1}) = u(t_i) + h \frac{du}{dt} |_{t_i} + \frac{h^2}{2!} \frac{d^2 u}{dt^2} |_{t_1 = \xi_1} \quad \xi_1 \in [t_i, t_{i+1}].$$  \hspace{1cm} (4.18)

This expression can be rearranged to provide an approximation for $\frac{du(t)}{dt}$. We have used the forward difference approximation in the preceding example, that is

$$u_{i+1} = u_i + h \frac{du_i}{dt}$$  \hspace{1cm} (4.19)

$$\frac{du_i}{dt} = \frac{u_{i+1} - u_i}{h}. \hspace{1cm} (4.20)$$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$u_i$</th>
<th>$u(t_i)$</th>
<th>$E(t_i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1.000</td>
<td>1.000</td>
<td>0</td>
</tr>
<tr>
<td>0.02</td>
<td>0.9820</td>
<td>0.9825</td>
<td>-0.0005</td>
</tr>
<tr>
<td>0.04</td>
<td>0.9650</td>
<td>0.9659</td>
<td>-0.0009</td>
</tr>
<tr>
<td>0.06</td>
<td>0.9489</td>
<td>0.9502</td>
<td>-0.0013</td>
</tr>
<tr>
<td>0.08</td>
<td>0.9336</td>
<td>0.9354</td>
<td>-0.0018</td>
</tr>
<tr>
<td>0.10</td>
<td>0.9191</td>
<td>0.92123</td>
<td>-0.0021</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.20</td>
<td>0.8402</td>
<td>0.8594</td>
<td>-0.0192</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.30</td>
<td>0.7862</td>
<td>0.8093</td>
<td>-0.0231</td>
</tr>
</tbody>
</table>

1 The symbol $\forall$ means ‘for all’.
It is important at this stage to realize that \( u(t_i) \) and \( u_i \) are different as are \( \frac{du}{dt} \bigg|_{t_i} \) and \( \frac{du_i}{dt} \). The value \( u(t_i) \) is the exact solution evaluated at the time \( t_i \). The value \( u_i \) is the value of the function as computed using a first-order difference approximation for Eq. (4.17). Thus, \( u_i \) inherently contains any errors that have evolved during the calculation of \( u_1, u_2, u_3, \ldots u_i \). As a result, the accumulated error in the solution can be determined at time \( t_{i+1} \) by subtracting \( u(t_{i+1}) \) from \( u_{i+1} \). Let us do the subtraction, that is subtract Eq. (4.18) from Eq. (4.19). We obtain

\[
u_{i+1} - u(t_{i+1}) = u_i - u(t_i) + h \frac{du}{dt} \bigg|_{t_i} - h \frac{du}{dt} \bigg|_{t_i} = h^2 \frac{d^2u}{dt^2} \bigg|_{t=t_i}
\]

or, using the information from Eq. (4.17) to replace the first derivative in this expression, we obtain

\[
u_{i+1} - u(t_{i+1}) = u_i - u(t_i) + h [f(t_i, u_i) - f(t_i, u(t_i))]
\]

\[
- h^2 \frac{d^2u}{dt^2} \bigg|_{t=t_i}
\]

(4.22)

Now, let us define the error terms as

\[
\varepsilon_i = u_i - u(t_i).
\]

Substitution of these definitions into Eq. (4.22) and defining

\[
E = - h^2 \frac{d^2u}{dt^2} \bigg|_{t=t_i}
\]

we obtain

\[
\varepsilon_{i+1} = \varepsilon_i + h [f(t_i, u_i) - f(t_i, u(t_i))] + E.
\]

(4.25)

We now require the use of the mean value theorem which states the following:

Given a function \( f(t) \) is continuous on the closed interval \([a, b]\), and \( a < b \), and that \( f(t) \) is differentiable on the open interval \((a, b)\), then there is a point \( c \) in \((a, b)\) such that

\[
\frac{df}{dt} \bigg|_{c} = \frac{f(b) - f(a)}{b - a}.
\]

(4.26)

In other words, there exists a point \( c \) in the interval \((a, b)\) such that the line joining the function values at the endpoints of the interval \([a, b]\) is parallel to the tangent at \( c \). (see Fig. 4.1).

Now, we employ the mean value theorem, Eq. (4.26), using \( b \equiv u(t_i) \) and \( a \equiv u_i \) to write

\[
f(t_i, u(t_i)) - f(t_i, u_i) = \frac{df}{du} \bigg|_{(t_i, u_i)} [u(t_i) - u_i], \quad \bar{u}_i \in [u_i, u(t_i)],
\]

(4.27)

or

\[
f(t_i, u(t_i)) - f(t_i, u_i) = - \frac{df}{du} \bigg|_{(t_i, \bar{u}_i)} \varepsilon_i.
\]

(4.28)
4.2. CONVERGENCE

Figure 4.1: Illustration of the mean value theorem. The line connecting the endpoints of the curve is parallel to a tangent of the curve at some point \( c \) in the interval \( a - b \).

Notice the appearance of the mean value \( \bar{u} \) in this expression.

Substitution of Eq. (4.28) into Eq. (4.25) gives

\[
\varepsilon_{i+1} = \varepsilon_i + h \frac{d}{du}{\bigg|}_{(t_i, \varepsilon_i)} \varepsilon_i + E. \tag{4.29}
\]

Let

\[
p = \frac{df}{du}{\bigg|}_{(t_i, \varepsilon_i)} \tag{4.30}
\]

such that

\[
\varepsilon_{i+1} = \varepsilon_i + hp \varepsilon_i + E. \tag{4.31}
\]

To see how the error evolves over time, we now make a series of calculations while advancing through time using the following relationship:

\[
\varepsilon_{i+1} = \varepsilon_i (1 + hp) + E. \tag{4.32}
\]

This will allow us to see what happens to \( \varepsilon_{i+1} \) as we proceed step-wise from the initial conditions. Here, we are considering the special case of \( p \) and \( E \) constant (which they are not).

We begin by using the observation that when

\[
i = 0, \tag{4.33}
\]

\[
u_{i=0} = u(t_0) = u_0. \tag{4.34}
\]

Then, by successive substitution, we obtain using Eq. (4.32) the following sequence,
\[ \varepsilon_0 = 0 \quad (\text{since } u_{i=0} - u(t_0) = 0) \]  
\[ \varepsilon_1 = E \]  
\[ \varepsilon_2 = E [1 + hp] + E 
= E [2 + hp] \]  
\[ \varepsilon_3 = E [2 + hp] [1 + hp] + E 
= 3E + 3Ehp + Eh^2p^2 
= 3E [1 + hp] + Eh^2p^2, \]  
\[ \varepsilon_4 = [3E [1 + hp] + Eh^2p^2] [1 + hp] + E 
= 4E + 6Ehp + 4Eh^2p^2 + Eh^3p^3, \]  
\[ \varepsilon_{i+1} = (i + 1) E + O(Eh), \]

where all the terms of power higher than \( h \) are smaller than \( O(Eh) \) and are omitted.

Now, recall from Eq. (4.24) that
\[ E = O(h^2) \]  
and that
\[ t_{i+1} = t_0 + (i + 1) h. \]

We can now introduce Eqs. (4.41) and (4.42) into Eq. (4.40) to give an expression for \( \varepsilon_{i+1} \)

\[ \varepsilon_{i+1} \approx \frac{t_{i+1} - t_0}{h} E + O(Eh) \]
\[ \approx \frac{t_{i+1} - t_0}{h} O(h^2) + O(h^3) \]
\[ \approx O(h) + .... \] (4.43)

The above development shows that the total error is one order lower than the local error (that is the error per time step). Notice that as \( h \) goes to zero, \( \varepsilon_{i+1} \) also goes to zero, thereby illustrating convergence for the Euler Forward Integration method.

### 4.3 Consistency

We will now look at the issue of consistency which, simply stated, is the requirement that the approximating equation approach the original differential equation as the increment \( h \) approaches zero. Let us consider an example. Given the equation
\[ u(t_{i+1}) = u(t_i) + hf(t_i, u_i) + \frac{h^2}{2!} \frac{d^2u(t)}{dt^2} \bigg|_{t=\xi} \]  
or, dividing through by \( h \), we get
\[ \frac{u(t_{i+1}) - u(t_i)}{h} = f(t_i, u_i) + \frac{h}{2!} \frac{d^2u(t)}{dt^2} \bigg|_{t=\xi} \frac{E}{h} \] (4.45)
then if $E/h$ vanishes as $h \to 0$, (which it will do because $E/h = O(h^2)$ so $h$ appears only in the numerator of $E/h$), we say the difference equation is **consistent** with the differential equation, viz.

$$
\lim_{h \to 0} \left( \frac{u(t_{i+1}) - u(t_i)}{h} - f(t_i, u_i) - \frac{h}{2!} \frac{d^2 u(x)}{dx^2} \bigg|_{x = \xi} \right) = \frac{du(t_i)}{dt} - f(t_i, u). \quad (4.46)
$$

Thus, in Eq. (4.46) we have shown that Euler’s method is **consistent**, that is the difference equation becomes, in the limit as $h \to 0$, equal to the original differential equation.

### 4.4 Stability

The last property to consider is stability. It determines whether, as one advances through time, the numerical solution behaves as expected (stable) or alternatively generates an ever-increasing error due to round off in the computer (unstable). Clearly, an unstable difference equation is of no practical importance.

As an example, consider the approximation to the following equation

$$
\frac{du}{dt} = \alpha u \quad (4.47)
$$
given by

$$
u_{i+1} - u_i = \alpha h u_i, \quad t \in [0, b] \quad [h \equiv t_{i+1} - t_i]. \quad (4.48)
$$

Assume an exact solution, free of round-off error, is given by $\tilde{u}_i$. Then $\tilde{u}$ satisfies

$$
\tilde{u}_{i+1} - \tilde{u}_i = ah \tilde{u}_i. \quad (4.49)
$$

Now, subtract $\tilde{u}_i$ from $u_i$ in Eq. (4.48) to give

$$
u_{i+1} - \tilde{u}_{i+1} - (u_i - \tilde{u}_i) = \alpha h (u_i - \tilde{u}_i). \quad (4.50)
$$

Let

$$
\tilde{e}_i \equiv u_i - \tilde{u}_i \quad (4.51)
$$
be the round-off error committed at $t_i$. If $\tilde{e}_i$ increases as $i$ increases, we have a problem. Let us take a look at what happens to $\tilde{e}_i$ as $i$ increases. To do this we first write Eq. (4.50) as

$$
\tilde{e}_{i+1} - \tilde{e}_i = ah \tilde{e}_i \quad (4.52)
$$
or

$$
\tilde{e}_{i+1} = (1 + \alpha h) \tilde{e}_i. \quad (4.53)
$$

Now, consider the cumulative effect of this error as we proceed through time:

$$
\tilde{e}_{i+1} = (1 + \alpha h) \tilde{e}_i, \quad (4.54)
$$

$$
\tilde{e}_{i+2} = (1 + \alpha h) \tilde{e}_{i+1} = (1 + \alpha h)^2 \tilde{e}_i, \quad (4.55)
$$
\[ \tilde{e}_{i+3} = (1 + \alpha h) \tilde{e}_{i+2} = (1 + \alpha h)^2 \tilde{e}_{i+1} = (1 + \alpha h)^3 \tilde{e}_{i+1}, \quad (4.56) \]

and in general

\[ \tilde{e}_{i+n} = (1 + \alpha h)^n \tilde{e}_i. \quad (4.57) \]

Examination of Eq. (4.57) reveals that to assure no exponential error growth, the following inequality has to hold,

\[ |1 + \alpha h| \leq 1. \quad (4.58) \]

This we consider to be a stable approximation, conditional on the requirement expressed in Eq. (4.58).

### 4.4.1 Example of Stability

To further illustrate the concept of stability, we will consider the following example which assumes \( \alpha = -1 \) in Eq. (4.58). We have

\[ |1 + \alpha h| = |1 + (-1) h| = |1 - h|. \quad (4.59) \]

For

\[ |1 - h| \leq 1 \quad (4.60) \]

we deduce that

\[ -1 \leq (1 - h) \leq 1 \quad (4.61) \]

which requires

\[ h \leq 2, \quad h \geq 0. \quad (4.62) \]

Thus, the approximation of Eq. (4.57) is conditionally stable, that is, it is stable provided \( h \leq 2 \) (since \( h \) must always be greater than 0).

### 4.5 Lax Equivalence Theorem

To determine convergence using the above strategy, one needs to know the solution of the equations beforehand, but the reason we are trying to find the solution numerically is because the exact solution is not, in fact, generally known. Fortunately, there is a theorem that helps; it is the Lax Equivalence Theorem and it states ([Lax and Richmyer, 1956]):

**Theorem 3** Given a properly posed initial value problem and a finite-difference approximation to it that satisfies the consistency condition, stability is the necessary and sufficient condition for convergence.

This is a very important result from a practical perspective because it states that, if we can determine consistency and stability, which do not require knowledge of the exact solution, we can establish convergence.

For the sake of completeness we are now going to provide some very well known, very specific methodology for solving initial value problems that are widely used, these are known as the Runge–Kutta formulae.
4.6 Runge–Kutta Type Formulae

4.6.1 General Form

To illustrate the concepts behind the Runge–Kutta methods, we will consider the equation of interest to be
\[ \frac{du}{dt} = f(u, t). \]  
(4.63)

The general form of the Runge–Kutta type formula for this equation using a two-term expansion is
\[ u_{i+1} = u_i + ak_1 + bk_2 \]  
(4.64)

where
\[ k_1 = hf(u_i, t_i) \]  
(4.65)
\[ k_2 = hf(u_i + \beta k_1, t_i + \alpha h). \]  
(4.66)

4.6.2 Runge–Kutta First-Order Form (Euler’s Method)

In Eq. (4.64) let \( a = 1 \) and \( b = 0 \). Substitution of these values into Eq. (4.64) we get
\[ u_{i+1} = u_i + hf(u_i, t_i) \]  
(4.67)
or, rewriting this expression
\[ \frac{u_{i+1} - u_i}{h} = f(u_i, t_i). \]  
(4.68)

This is the forward finite difference approximation we derived earlier and denoted as the Euler method.

4.6.3 Runge–Kutta Second-Order Form

Now, consider the extension of the above method to a two–step scheme. Consider once again the equation
\[ \frac{du}{dt} = f(u, t). \]  
(4.69)

Let the coefficients in Eq. (4.64) take on the values
\[ a = b = 1/2 \quad \alpha = \beta = 1. \]  
(4.70)

Using these values in Eqs. (4.64–4.66) we obtain
\[ u_{i+1} = u_i + \frac{1}{2}k_1 + \frac{1}{2}k_2 \]  
(4.71)

where
\[ k_1 = hf(u_i, t_i) \]  
(4.72)
as before and
\[ k_2 = hf(u_i + k_1, t_i + h) \]  
(4.73)

which, when substituted into Eq. (4.71), gives
\[ u_{i+1} = u_i + \frac{1}{2} [hf(u_i, t_i) + hf(u_i + hf(u_i, t_i), t_i + h)]. \]  
(4.74)
Example of Runge–Kutta Second-Order Form

Let us now consider an illustrative example in which we consider the equation
\[
\frac{du}{dt} = \tan(u) + 1, \quad t \in [1, 1.075],
\] (4.75)

with initial condition
\[
u_0 = 1,
\] (4.76)

and \(h = 0.025\). Since the interval is of length 0.075, there will be three time steps. The equation to be used is Eq. (4.71). We will proceed by calculating \(k_1\) and \(k_2\) and then substitute these into Eq. (4.71). With \(u\) in radians we obtain for \(k_1\) and \(k_2\)

\[
k_1 = hf(u_0, t_0)
= 0.025(\tan(1) + 1)
= 0.025(2.56)
= 0.064,
\] (4.77)

\[
k_2 = h[\tan(u_0 + k_1, t_0 + h) + 1]
= h[\tan((1 + 0.064), 1.025) + 1]
= 0.025[\tan(1.064) + 1]
= 0.070.
\] (4.78)

Substitution of these values into Eq. (4.71), we obtain

\[
u_1 = u_0 + \frac{1}{2}k_1 + \frac{1}{2}k_2
= 1 + \frac{1}{2}0.064 + \frac{1}{2}0.070
= 1.067.
\] (4.79)

Now, for step two, we obtain

\[
k_1 = hf(u_1, t_1)
= 0.025[\tan(1.067) + 1]
= 0.025(1.81 + 1)
= 0.070,
\] (4.80)

\[
k_2 = h[\tan(u_1 + k_1, t_1 + h) + 1]
= h[\tan((1.067 + 0.070), 1.050) + 1]
= 0.025[(\tan(1.137) + 1]
= 0.079,
\] (4.81)
4.6. RUNGE–KUTTA TYPE FORMULAE

\[ u_2 = u_1 + \frac{1}{2}k_1 + \frac{1}{2}k_2 \]
\[ = 1.067 + \frac{1}{2}0.070 + \frac{1}{2}0.79 \]
\[ = 1.14. \]  
\( (4.82) \)

And finally, for step three, we have

\[ k_1 = hf (u_2, t_2) \]
\[ = 0.025[\tan (1.14) + 1] \]
\[ = 0.025(2.17 + 1) \]
\[ = 0.0794, \]  
\( (4.83) \)

\[ k_2 = h[\tan(u_2 + k_1, t_2 + h) + 1] \]
\[ = h[\tan((1.14 + 0.0794), 1.075) + 1] \]
\[ 0.025[(\tan(1.22) + 1] \]
\[ = .0933, \]  
\( (4.84) \)

\[ u_3 = u_2 + \frac{1}{2}k_1 + \frac{1}{2}k_2 \]
\[ = 1.14 + \frac{1}{2}0.0794 + \frac{1}{2}0.0933 \]
\[ = 1.23. \]  
\( (4.85) \)

The analytic solution to this problem is

\[ 0.5 (u + \ln |\sin x + \cos x|) + 0.338 - t = 0 \]  
\( (4.86) \)

and for \( t = 1.075 \), we obtain \( u = 1.23 \) so the solution is exact to three significant figures.

**Geometric Interpretation**

We will now use Fig. 4.2 to look at one way to explain the difference between Euler’s method and the Runge–Kutta method. The Euler method uses the equation

\[ u_{i+1} = u_i + hf (u_i, t_i) \]  
\( (4.87) \)

which, using the original equation Eq. (4.17) is equivalent to

\[ u_{i+1} = u_i + h \frac{du}{dt} \bigg|_{t_i}. \]  
\( (4.88) \)

Thus, the Euler method projects from \( u_i \) to \( u_{i+1} \) as represented by the curve \( P_1 \).

The Runge–Kutta method uses

\[ u_{i+1} = u_i + \frac{h f (u_i, t_i)}{2} + \underbrace{hf (u_i + hf (u_i, t_i), t_i + h)}_{P_2} \]
\[ + \underbrace{hf (u_i + hf (u_i, t_i), t_i + h)}_{P_3} \]  
\( (4.89) \)
which we can interpret as follows: the $P_1$ term is the equivalent of the Euler method. The term $P_2$ is equivalent to

$$P_2 = h \frac{d}{dt} \left( u_i + h \frac{du}{dt}\bigg|_{t_i}, t_i + h \right) = h \frac{du}{dt}\bigg|_{i+1}$$

(4.90)

which is $\frac{du}{dt}$ evaluated at the end of the interval, that is at $t = t_0 + (i + 1)h$. The Runge–Kutta method uses the average of $P_1$ and $P_2$ to yield

$$P_3 = \frac{P_1 + P_2}{2}$$

(4.91)

which are the slopes at $t_i$ and $t_{i+1}$. Thus, the Runge–Kutta method uses the arithmetic mean value of the derivatives evaluated at the end points of the interval.

### 4.7 Chapter Summary

The solution of initial value problems using numerical approximations is the focus of this chapter. The three fundamental properties of a numerical approximation, namely consistency, convergence, and stability are presented and examples are provided to assist in understanding these concepts. Euler’s classical method along with Runge–Kutta formulations is also described.

### 4.8 Problems

1. Consider the equation

$$\frac{du}{dt} + u = 0.$$  

Since this is a first-order equation in time we need an initial condition. Assume the initial condition is

$$u(0) = 1.$$
Further assume that you are going to solve this equation using the forward difference scheme
\[
\frac{u_{i+1} - u_i}{h} + u_i = 0.
\]
and you are using a value of \(h = 3\). Calculate the first five time steps and observe whether the approximation is stable or unstable. Now, do the same calculation using \(h = 3/2\) and make the same observation. How is this result consistent with the theory presented in this chapter?

2. You find it necessary to forecast the rate of heating of a widget. The governing equation is

\[
\frac{dT}{dt} - \beta T = 0 \quad t \in (0, \infty), \quad (\beta \text{ is real})
\]

where \(\beta\) is an empirical constant and is positive. You decide to use the difference formula:

\[
\frac{T_{i+1} - T_i}{h} - \beta T_{i+1} = 0
\]

where \(h = t_i - t_{i-1}\) and \(t_i = ih\). Show how you determine the stability of this scheme and describe what the stability constraints are.

Include the following steps:

1. (a) Define your error term and represent it as \(e_i\);
   (b) Substitute the error into Eq. (4.7);
   (c) Show how the error evolves over each time step;
   (d) Determine the requirements on \(\beta\) and \(\Delta t\) for stability.

3. What is the difference between the error of the approximation and the round-off error?

4. Will the number of significant digits in the representation of a number affect the stability of an otherwise unstable scheme. In other words, can you use a computer which represents numbers with more bits to overcome stability problems?

5. In the truncation errors involving \(h\), for example, in Eq. (4.3), it is assumed that \(h \leq 1\) so that as the power of \(h\) increases, its magnitude decreases. What happens when \(h > 1\)? How then do you demonstrate consistency or establish the error of the approximation?
Bibliography


Chapter 5

Weighted Residuals Methods

We are now going to discuss the concept of the method of weighted residuals which, along with the interpolation theory presented above, will allow us to create an array of different numerical techniques. To start the discussion we consider the equation

\[ \frac{du}{dx} = f(x) \quad x \in [a, b]. \]  

(5.1)

From Lagrange interpolation (see Section 1.8, page 14), we can write the approximation to \( u(x) \) as

\[ u(x) = \sum_{j=0}^{n} \ell_j(x) u_j + E(x) \]  

(5.2)

or

\[ u(x) = \hat{u}(x) + E(x). \]  

(5.3)

where

\[ \hat{u}(x) = \sum_{j=0}^{n} \ell_j(x) u_j \]  

(5.4)

and \( \ell_j(x) \) is assumed to be a linear Lagrange polynomial with \( n + 1 \) nodes on the interval \( x \in [a, b] \). Substitution of Eq. (5.3) into Eq. (5.1) gives

\[ \frac{d\hat{u}(x)}{dx} - f(x) = - \frac{dE(x)}{dx} = R(x) \]  

(5.5)

where \( R(x) \neq 0 \) and is defined as a residual error or just a residual. The residual is the difference between the value of the differential equation using the exact solution \( u(x) \) and the value of the differential equation when we use the approximate solution \( \hat{u}(x) \).

Let \( w_i(x), i = 0, \ldots, n \) be a set of weighting functions. At this point, we have not specified what they might be, but we assume they exist. Next, multiply the residual \( R(x) \) by each of the weighting functions \( w_i(x) \), and integrate each of the resulting products over the domain of interest. Now, for each weighting function \( w_i(x), i = 0 \ldots n \), set the result to zero. Using mathematical notation we have

\[ \int_x R(x) w_i(x) \, dx = 0 \quad i = 0, 1, 2, \ldots, n \quad x \in [x_0 = a, x_n = b]. \]  

(5.6)
This weighted residual relationship is sometimes written using inner-product notation as

\[ \int_x R(x) w_i(x) \, dx = \langle R(x), w_i(x) \rangle = 0 \quad i = 0, 1, 2, \ldots, n \quad x \in [x_0 = a, x_n = b] \]  

(5.7)

where \( \langle \cdot \rangle \) denotes the inner product.

Next substitute Eq. (5.5) into (5.6) to get

\[ \int_x \left( \frac{d\hat{u}(x)}{dx} - f(x) \right) w_i(x) \, dx = 0 \quad i = 0, 1, 2, \ldots, n. \]  

(5.8)

If we introduce the definition of \( \hat{u}(x) \) found in Eq. (5.2) into Eq. (5.8) we get

\[ \int_x \left( \frac{d}{dx} \sum_{j=0}^{n} \ell_j(x) u_j - f(x) \right) w_i(x) \, dx = 0 \quad i = 0, 1, 2, \ldots, n. \]  

(5.9)

Since the variables \( u_j \) are simply numbers, we can take them out of the integration and obtain

\[ \sum_{j=0}^{n} u_j \int_x \frac{d}{dx} \ell_j(x) w_i(x) \, dx - \int_x f(x) w_i(x) \, dx = 0 \quad i = 0, 1, 2, \ldots, n. \]  

(5.10)

Examination of Eq. (5.10) reveals that we have \( n + 1 \) equations, one for each value of \( i = 0, 1, 2, \ldots, n \). Each equation has \( n + 1 \) unknowns, \( u_j(x) \), \( j = 0, 1, \ldots, n \). So, at least in theory, once we select the form of the functions \( w_i(x) \), we should be able to solve for the \( n + 1 \) values of \( u_j \). Different Lagrange polynomials \( \ell_j(x) \) and different weighting functions \( w_i(x) \) will give different numerical methods.

Once we have calculated the \( u_j \) values we can use Eq. (5.4) to obtain \( \hat{u}(x) \) anywhere in the interval \( x \in [a, b] \). We will consider a number of different forms for the weighting functions and thereby derive a corresponding number of numerical approaches. We will start with the ‘finite volume’ or ‘subdomain’ method.

### 5.1 Finite Volume or Subdomain Method

To obtain the finite volume method (also known as the subdomain method) we select as our weighting functions

\[ w_i(x) = \begin{cases} 1, & x \in \left[ x_i - \frac{h}{2}, x_i + \frac{h}{2} \right] \\ 0, & x \notin \left[ x_i - \frac{h}{2}, x_i + \frac{h}{2} \right] \end{cases} \]  

(5.11)

which are illustrated in Fig. 5.1. In terms of the local coordinate system, that is the one tied to the element, we have

\[ w_0(\chi) = \begin{cases} 1, & \chi \in [0, 0.5] \\ 0, & \chi \in [0.5, 1.0] \end{cases} \]  

(5.12)

and

\[ w_1(\chi) = \begin{cases} 1, & \chi \in [0.5, 1.0] \\ 0, & \chi \in [0, 0.5] \end{cases} \]  

(5.13)
Next, we decide on the polynomial functions which we wish to use in approximating $u(x)$. We will call these approximating functions basis functions. As noted earlier, we will initially use the linear Lagrange polynomials shown in Fig. 1.10 on page 18 as our basis functions. In the following representation of our approximating function $\hat{u}(x)$

$$\hat{u}(x) = \sum_{j=0}^{1} \ell_j(\chi) u_j \quad \chi \in [0, 1].$$

The linear Lagrange polynomials are given by

$$\ell_0(\chi) = \frac{\chi_1 - \chi}{\chi_1 - \chi_0}$$

and

$$\ell_1(\chi) = \frac{\chi - \chi_0}{\chi_1 - \chi_0}.$$  

The $x$ and $\chi$ coordinates are defined by Eq. (1.89) on page 16 and in Fig. 1.10. The combined weighting and basis functions are shown in Fig. 5.2. Keep in mind that we are using two terms to describe the discretization of the $x$ axis. On the one hand, we have the finite element, this is the distance between any two nodes in a system characterized by linear basis functions. On the other hand we have the finite volume which is characterized by the segment of the $x$ axis under the finite volume. So, in Fig. 5.2 an element $e_i$ would be characterized by the segment from node $x_{i-1}$ to node $x_i$ and the finite volume associated with the node at $x_i$ would be the segment defined by $\frac{x_i - x_0}{2} + \frac{x_{i+1} - x_1}{2}$.  

Figure 5.1: Representation of the finite volume weighting function.

we replace $u(x_j)$ with simply $u_j$ to simplify notation. With this substitution we obtain for the approximation defined in the local coordinates $\chi$
5.1.1 Example

In this subsection, we will consider an example of the application of the finite volume approximation to the equation

$$\frac{du(x)}{dx} + u(x) = 1, \quad x \in [0, 1]$$  \hspace{1cm} (5.17)

$$u(0) = u_0 = 2.$$  \hspace{1cm} (5.18)

The analytic solution for this problem is

$$u(x) = 1 + \exp(-x).$$  \hspace{1cm} (5.19)

We first write the residual approximation

$$\frac{d\hat{u}(x)}{dx} + \hat{u}(x) - 1 = R(x).$$  \hspace{1cm} (5.20)

Next, we write our linear Lagrange basis functions for two elements (3 nodes). Each element will be of length $e_i = 0.5$, $i = 1, 2$ as will the middle volume. The volumes at each end of the domain are smaller because they are truncated by the existence of the boundaries.

From the method of weighted residuals (Eq. (5.6)) we have (see Fig. 5.3):

$$\int_{0}^{1.0} R(x) w_i(x) \, dx = 0 \quad i = 0, 1, 2.$$  \hspace{1cm} (5.21)

Let us examine the weighting function $w_0(x)$; it has the value of 1 only over one-half of element 1 and 0 elsewhere. Thus, since the integrand is defined only over the first half of the first element, the integration needs to be performed only over the first half of the first element, that is.
Figure 5.3: Weighting function and basis function for finite volume method example. In the example discussed in this section the subscripts \( i - 1 = 0 \), \( i = 1 \), and \( i + 1 = 2 \).

\[
\int_0^{1.0} R(x) w_0(x) \, dx = \int_0^{0.25} R(x) \, dx = 0.
\]  

(5.22)

For weighting function \( w_1(x) \), we have non-zero values of this function over half of the first element and half of the second as shown in Figs. 5.1 and 5.3. Thus integration is required over this interval that spans parts of two elements, that is

\[
\int_0^{1.0} R(x) w_1(x) \, dx = \int_0^{0.5} R(x) \, dx + \int_0^{0.75} R(x) \, dx = 0.
\]  

(5.23)

Finally, for \( w_2(x) \) we have integration over half of element 2 only, that is

\[
\int_0^{1.0} R(x) w_2(x) \, dx = \int_{0.75}^{1.0} R(x) \, dx = 0.
\]  

(5.24)

We now perform the integrations, one element at a time. Note that each complete weighting function generates one equation. This means that the equation for one
weighting function will contain integrals from more than one element. In the case of a one-dimensional problem such as we are considering here, there will be two element integrals for each weighting function. In two dimensions we will see that there are several element integrations per weighting function.

Because the first weighting function \( w_0(x) \) occupies only the first element, a complete equation is therefore obtained by integration only over the first element. The appropriate integration therefore is

\[
\int_0^{0.25} R(x) \, dx = \int_0^{0.25} \left( \frac{d\hat{u}}{dx} + \hat{u} - 1 \right) \, dx
\]

\[
= \int_0^{0.25} \left[ \left( \frac{d\ell_0}{dx} u_0 + \ell_0 u_0 \right) + \left( \frac{d\ell_1}{dx} u_1 + \ell_1 u_1 \right) - 1 \right] \, dx
\]

\[
= \int_0^{0.25} \left[ u_0 \left( \frac{d\ell_0}{dx} + \ell_0 \right) + u_1 \left( \frac{d\ell_1}{dx} + \ell_1 \right) - 1 \right] \, dx
\]

\[
= \int_0^{0.25} \left[ u_0 \left( \frac{-1}{h} + \frac{h - x}{h} \right) + u_1 \left( \frac{1}{h} + \frac{x}{h} \right) - 1 \right] \, dx
\]

\[
= u_0 \left( -5/16 \right) + u_1 \left( 9/16 \right) - 1/4 \]

\[
= 0. \tag{5.25}
\]

or

\[
u_0 \left( -5/16 \right) + u_1 \left( 9/16 \right) - 1/4 = 0.
\]

This is our first finite volume equation.

Although it is not really necessary for a one-dimensional problem such as we are considering here, in order to get accustomed to using the local coordinate system \( \chi \), we now consider the above integration using \( \chi \). The first step is to transform Eq. (5.22) into a form that employs \( \chi \). Since \( \chi \) is a function of \( x \), we can write \( \hat{u}(x) \) in terms of \( \chi \) to give \( \hat{u}(\chi) \)

\[
\int_0^{0.5} R(\chi) \frac{dx}{d\chi} d\chi = \int_0^{0.5} \left( \frac{d\hat{u}(\chi)}{d\chi} \frac{dx}{d\chi} + \hat{u}(\chi) - 1 \right) \frac{dx}{d\chi} d\chi \tag{5.26}
\]

where we recall that now the value of \( h \) is 1.0. The next challenge is obtaining \( \frac{dx}{d\chi} \). We know from Eq. (1.89) on page 16, that is

\[
x = \frac{\chi_1 - \chi}{\chi_1 - \chi_0} x_1 + \frac{\chi - \chi_0}{\chi_1 - \chi_0} x_2 \tag{5.27}
\]

so we can differentiate \( x \) to get, after rearranging,

\[
\frac{dx}{d\chi} = \frac{x_2 - x_1}{\chi_1 - \chi_0}
\]

\[
= \frac{h}{1}. \tag{5.28}
\]

Similarly, \( \frac{dx}{d\chi} = 1/h \). Thus, Eq. (5.26) becomes (using the definitions provided by Eqs. (5.15) and (5.16)
\[\int_{0}^{0.25} R(x) \, dx = \int_{0}^{0.5} \left( \frac{d\hat{u}(\chi)}{d\chi} + \hat{u}(\chi) - 1 \right) \, dx \]
\[= \int_{0}^{0.5} \left( \frac{d\hat{u}(\chi)}{d\chi} \frac{1}{h} + \hat{u}(\chi) - 1 \right) \, h \, d\chi \]
\[= \int_{0}^{0.5} \left[ \left( \frac{1}{h} \frac{d\ell_{0}(\chi)}{d\chi} u_{0} + \ell_{0}(\chi) u_{0} \right) + \left( \frac{1}{h} \frac{d\ell_{1}(\chi)}{d\chi} + \ell_{1}(\chi) u_{1} \right) - 1 \right] \, h \, d\chi \]
\[= \int_{0}^{0.5} \left[ u_{0} \left( \frac{1}{h} \left( -\frac{1}{2} \right) + 1 - \chi \right) + u_{1} \left( \frac{1}{h} \left( \frac{1}{2} \right) + \frac{\chi}{1} \right) - 1 \right] \, h \, d\chi \]
\[= \left[ u_{0} \left( -\frac{7}{16} + 1/2 - 1/4 \right) + u_{1} \left( \frac{7}{16} + 1/4 \right) - 1/2 \right] \frac{1}{2} \]
\[= u_{0} (-5/16) + u_{1} (9/16) - 1/4 \]
\[= 0 \quad (5.29)\]

which is the same result as we obtained in Eq. (5.25). Notice the change in the limits of integration due to the fact that we are now integrating over \( \chi \) rather than \( x \).

For the second weighting function \( w_{1}(x) \), two elements are involved. For element 1 we have for the global system

\[\int_{0.25}^{0.5} R(x) \, dx = \int_{0.25}^{0.5} \left( \frac{d\hat{u}(x)}{dx} + \hat{u}(x) - 1 \right) \, dx \]
\[= \int_{0.25}^{0.5} \left[ u_{0} \left( \frac{d\ell_{0}(x)}{dx} + \ell_{0}(x) \right) + u_{1} \left( \frac{d\ell_{1}(x)}{dx} + \ell_{1}(x) \right) - 1 \right] \, dx \]
\[= \int_{0.25}^{0.5} \left[ u_{0} \left( -\frac{7}{16} + \frac{h-x}{h} \right) + u_{1} \left( \frac{7}{16} + \frac{x}{h} \right) - 1 \right] \, dx \]
\[= u_{0} \left( -\frac{7}{16} \right) + u_{1} \left( \frac{11}{16} \right) - \frac{1}{4} \quad (5.30)\]

and for the local system (using the definitions of \( \ell_{0}(\chi) \) and \( \ell_{1}(\chi) \) given in Eqs. (1.91) and (1.92) on page 17)
\int_{0.25}^{0.50} R(x) \, dx = \int_{0.5}^{1.0} \left( \frac{du(\chi)}{d\chi} \frac{dx}{d\chi} + \hat{u}(\chi) - 1 \right) \frac{dx}{d\chi} d\chi \\
= \int_{0.5}^{1.0} \left( \frac{du(\chi)}{d\chi} \frac{1}{h} + \hat{u}(\chi) - 1 \right) h d\chi \\
= \int_{0.5}^{1.0} \left[ \left( \frac{1}{h} \frac{dG(\chi)}{d\chi} w_0 + \ell_0(\chi) w_0 \right) + \left( \frac{1}{h} \frac{dG_1(\chi)}{d\chi} w_1 + \ell_1(\chi) w_1 \right) - 1 \right] h d\chi \\
= \int_{0.5}^{1.0} \left[ w_0 \left( \frac{1}{h} \frac{dG(\chi)}{d\chi} + \ell_0(\chi) \right) + w_1 \left( \frac{1}{h} \frac{dG_1(\chi)}{d\chi} + \ell_1(\chi) \right) - 1 \right] h d\chi \\
= \int_{0.5}^{0.10} \left\{ w_0 \left[ \left( \frac{1}{h} \frac{1}{1} + \frac{1 - \chi}{1} \right) + w_1 \left[ \left( \frac{1}{h} \frac{1}{1} + \chi \right) \right] - 1 \right] \right\} h d\chi \\
= \left[ w_0 \left( \frac{-\chi}{h} + \chi - \frac{\chi^2}{2} \right) \right]_{0.5}^{1.0} + w_1 \left[ \left( \frac{\chi}{h} + \frac{\chi^2}{2} \right) \right]_{0.5}^{1.0} \right] h \\
= \left[ w_0 \left( \left[ -\frac{7}{16} + 1 - \frac{1}{2} \right] - \left( -\frac{1}{2h} + \frac{1}{2} - \frac{1}{8} \right) \right) + w_1 \left[ \left( \frac{1}{h} + \frac{1}{2} \right) - \left( \frac{1}{2h} + \frac{1}{8} \right) \right] - 1/2 \right] \right] 1/2 \\
= w_0 \left( -\frac{7}{16} \right) + w_1 \left( \frac{11}{16} \right) - 1/4. \tag{5.31}

For element 2, we note that at the local level the integrals are the same as for the case shown in Eq. (5.29), although the values of the coefficients are now \( u_1 \) and \( u_2 \).

We now combine the information regarding \( w_1(x) \) from both element 1 and element 2. We need both of these integrals to give us the complete integral for \( w_1(x) \) and we need the complete integral to form the equation for \( w_1(x) \) Thus, combining the integral in Eq. (5.29) and (5.31), we get

\[
\int_{0.25}^{0.50} R(x) \, dx + \int_{0.5}^{0.75} R(x) \, dx = \int_{0}^{1.0} R(\chi) \, d\chi \\
= \int_{0}^{1.0} \left( \frac{d\hat{u}(\chi)}{d\chi} \frac{dx}{d\chi} + \hat{u}(\chi) - 1 \right) \frac{dx}{d\chi} d\chi + \int_{0}^{0.5} \left( \frac{d\hat{u}(\chi)}{d\chi} \frac{dx}{d\chi} + \hat{u}(\chi) - 1 \right) \frac{dx}{d\chi} d\chi \\
= \left[ w_0 \left( -\frac{7}{16} \right) + u_1 \left( \frac{11}{16} \right) - 1/4 \right]_{e=1} + \left[ u_1 \left( -\frac{5}{16} \right) + u_2 \left( \frac{9}{16} \right) - 1/4 \right]_{e=2}. \tag{5.32}
\]

The two integrals in Eq. (5.32) constitute all the information associated with weighting function \( w_1(\chi) \). Thus, we can set their sum to zero and get a complete equation, that is
Let us write this as a matrix equation

\[
\int_{0.25}^{0.5} R(x) \, dx + \int_{0.5}^{0.75} R(x) \, dx
\]

\[
= [u_0 (-7/16) + u_1 (11/16) - 1/4]_{e=1} + [u_1 (-5/16) + u_2 (9/16) - 1/4]_{e=2}
\]

\[
= 0.
\] (5.33)

Now consider the last weighting function \( w_2(x) \). We have only one integral because \( w_2(x) \) only appears in the last element. We therefore obtain

\[
\int_{0.75}^{1.0} R(x) \, dx
\]

\[
= \int_{0.5}^{1.0} \left( \frac{d\hat{u}}{d\chi} + \hat{u} - 1 \right) \, d\chi
\]

\[
= \int_{0.5}^{1.0} \left( \frac{d\hat{u}}{d\chi} \hat{\chi} + \hat{u} (\chi) - 1 \right) \, d\chi
\]

\[
= \int_{0.5}^{1.0} \left[ \frac{1}{h} \frac{d\ell_0 (\chi)}{d\chi} u_0 + \ell_0 (\chi) u_0 \right] + \frac{1}{h} \frac{d\ell_1 (\chi)}{d\chi} u_1 + \ell_1 (\chi) u_1 - 1 \right] \, d\chi
\]

\[
= \int_{0.5}^{1.0} \left[ u_0 \left( \frac{1}{h} \frac{d\ell_0 (\chi)}{d\chi} + \ell_0 (\chi) \right) + u_1 \left( \frac{1}{h} \frac{d\ell_1 (\chi)}{d\chi} + \ell_1 (\chi) \right) - 1 \right] \, d\chi
\]

\[
= \left[ u_0 \left( -\chi + \chi^2 - \frac{1}{2} \right) + u_1 \left( \chi^2 + \frac{1}{2} \right) \right]_{0.5}^{1.0} \frac{1}{h} - \chi_{0.5}^{1.0} h
\]

\[
= \left\{ u_0 \left[ \left( \frac{1}{h} + 1 - \frac{1}{2} \right) - \frac{1}{2} \frac{1}{h} + \frac{1}{2} - \frac{1}{8} \right] + u_1 \left[ \left( \frac{1}{h} + 1 \right) - \left( \frac{1}{2h} + \frac{1}{8} \right) \right] - 1/2 \right\} 1/2
\]

\[
= u_1 (-7/16) + u_2 (11/16) - 1/4.
\] (5.34)

Since this is all the information we will obtain using \( w_2 \), we can write

\[
\int_{0.75}^{1.0} R(x) \, dx
\]

\[
= u_1 (-7/16) + u_2 (11/16) - 1/4
\]

\[
= 0.
\] (5.35)

We now have three equations, Eqs. (5.25), (5.33) and (5.35) in three unknowns, \( u_0, u_1, \) and \( u_2 \), which can be written as

\[
[u_0 (-5/16) + u_1 (9/16) - 1/4]_{e=1} = 0
\]

\[
[u_0 (-7/16) + u_1 (11/16) - 1/4]_{e=1} + [u_1 (-5/16) + u_2 (9/16) - 1/4]_{e=2} = 0
\]

\[
[u_1 (-7/16) + u_2 (11/16) - 1/4]_{e=2} = 0.
\] (5.36)

Let us write this as a matrix equation

\[
\begin{bmatrix}
-\frac{5}{16} & \frac{9}{16} & 0 \\
-\frac{7}{16} & \frac{9}{16} & \frac{5}{16} \\
0 & -\frac{7}{16} & \frac{5}{16}
\end{bmatrix}
\begin{bmatrix}
\begin{bmatrix}
u_0 \\
u_1 \\
u_2
\end{bmatrix}
= \begin{bmatrix}
\frac{1}{4} \\
\frac{1}{4} \\
\frac{1}{4}
\end{bmatrix}.
\end{bmatrix}
\] (5.37)
However, we have information on \( u_0 \) from Eq. (5.17), namely that
\[
 u(0) = u_0 = 2.
\]
Thus, we can replace \( u_0 \) with 2. Now, since we know \( u_0 \), we can multiply the first column by 2 and place the result on the right-hand side of the equation. The result is
\[
\begin{bmatrix}
 \frac{9}{16} & \frac{5}{16} & 0 \\
 -\frac{7}{16} & \frac{9}{16} & \frac{3}{16}
\end{bmatrix}
\begin{bmatrix}
 2 \\
 u_1 \\
 u_2
\end{bmatrix} = \begin{bmatrix}
 \frac{1}{4} + 2 \left( \frac{5}{16} \right) \\
 \frac{1}{2} + 2 \left( \frac{3}{16} \right)
\end{bmatrix}.
\]
(5.38)

Since we know the value of \( u_0 \), we do not need the first line of the matrix and so we remove it along with the first column, which is also no longer needed, to obtain
\[
\begin{bmatrix}
 -\frac{3}{16} & \frac{9}{16} \\
 \frac{3}{16} & \frac{11}{16}
\end{bmatrix}
\begin{bmatrix}
 u_1 \\
 u_2
\end{bmatrix} = \begin{bmatrix}
 \frac{11}{4} \\
 \frac{1}{4}
\end{bmatrix}
\]
(5.39)

from which we get
\[
\begin{bmatrix}
 u_1 \\
 u_2
\end{bmatrix} = \begin{bmatrix}
 1.60 \\
 1.38
\end{bmatrix}
\]
(5.40)

with these values of \( u_i(x) \) and \( u_0(x) \) we can write from the Lagrange expansion
\[
\hat{u}(x) = \ell_0(x) u_0 + \ell_1(x) u_1 + \ell_2(x) u_2
\]
\[
\hat{u}(x) = \ell_0(x) (2) + \ell_1(x) (1.60) + \ell_2(x) (1.38)
\]
(5.41)

where the \( x \)-dependent Lagrange may span two elements. Thus, we can obtain a value for the approximation \( \hat{u}(x) \) anywhere in the domain of interest. From the analytic solution, we have
\[
u(0.5) = u_1 = 1 + exp(-0.5) = 1 + 0.61 = 1.61
\]
(5.42)

and
\[
u(1.0) = u_2 = 1 + exp(-1.0) = 1 + 0.37 = 1.37.
\]
(5.43)

<table>
<thead>
<tr>
<th>Location</th>
<th>( u(x) )</th>
<th>( \hat{u}(x) )</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x = 0.5 )</td>
<td>1.61</td>
<td>1.60</td>
<td>0.01</td>
</tr>
<tr>
<td>( x = 1.0 )</td>
<td>1.37</td>
<td>1.38</td>
<td>0.01</td>
</tr>
</tbody>
</table>

The numerical and analytical solutions are remarkably close given only a three volume approximation was used.

While the above approach to setting up the matrix equation seems like an appropriate way to proceed, this is not the way the problem is solved in practice. Rather, each element is allotted its own \( 2 \times 2 \) matrix which is called the element coefficient matrix. The elements of the element coefficient matrix are the integrals identified with that element. To see how this works first refer to Eq. (5.36). Note that there is notation indicating from which element each coefficient is obtained. The information in the first row of this equation comes from element 1, the second row from elements 1 and 2 and the third row from element 2. In the following we collect the information from each element.

The element-defined terms in the equations must be (and have been) summed to form complete equations identified with a volume. For example to have all the information associated with volume weighting function \( w_1 \) we need to sum the information in the second row of the matrix associated with element 1 and the first row of element 2 in Eq. (5.44):
5.1. FINITE VOLUME OR SUBDOMAIN METHOD

Let us sum all the coefficients in both matrices that are associated with each unknown value of \( u_i \). We get

\[
\begin{bmatrix}
\begin{bmatrix}
\frac{-5}{16} & 9/16 \\
\text{Element 1} & 1
\end{bmatrix} & \begin{bmatrix}
\frac{9}{16} \\
\text{Element 2} & 2
\end{bmatrix}
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1
\end{bmatrix} =
\begin{bmatrix}
\frac{-5}{16} & 9/16 \\
\text{Element 1} & 1
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix}.
\]

(5.44)

which is identical to the left-hand side of Eq. (5.37). The \( 3 \times 3 \) matrix in Eq. (5.45) is called the global coefficient matrix. With this information, the equations can be formulated and solved as seen in Eqs. (5.37) through (5.40). In practice, each element coefficient matrix is generated and then the information is transferred from each one to the global matrix.

Recall that, we performed the integrations using the local \( \chi \) coordinate system. To do this, we needed to know the relationship between the location of the nodes in the global \( \chi \) and local \( x \) coordinate systems. This relationship is determined through the use of a table or map such as shown below. For each element, the correspondence between the locations of the node numbers in the global and local coordinates are provided:

<table>
<thead>
<tr>
<th>Element 1</th>
<th>Element 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x_0 )</td>
<td>( \chi_0 )</td>
</tr>
<tr>
<td>( x_1 )</td>
<td>( \chi_1 )</td>
</tr>
</tbody>
</table>

(5.46)

5.1.2 Finite Difference Interpretation of the Finite Volume Method

Consider an interval \( x \in [x_a = 0, x_b = 2h] \). (see Fig. 5.4). Then the integrals appearing in the preceding example (Section 5.1.1) for the case of \( w_1 \) are

\[
\int_0^{2h} R(x) w_1(x) dx = \int_0^{2h} \left( \frac{d\hat{u}(x)}{dx} + \hat{u}(x) - 1 \right) w_1(x) dx = 0
\]

(5.47)

\[
\int_0^{2h} \frac{d\hat{u}(x)}{dx} w_1(x) dx = \int_{h/2}^{(3/2)h} \frac{d\hat{u}(x)}{dx} dx = \frac{1}{2} (u_2 - u_0)
\]

(5.48)

\[
\int_0^{2h} \hat{u}(x) w_1(x) dx = \int_{h/2}^{(3/2)h} \hat{u}(x) dx = \left( \frac{1}{8} u_0 + \frac{3}{4} u_1 + \frac{1}{8} u_2 \right) h
\]

(5.49)

\[
\int_0^{2h} (1) w_1(x) dx = \int_{h/2}^{(3/2)h} (1) dx = (1) h.
\]

(5.50)

Combining the integrals, we get

\[
\int_0^{2h} \left( \frac{d\hat{u}(x)}{dx} + \hat{u}(x) - 1 \right) w_1(x) dx = \frac{1}{2} (u_2 - u_0) + \left( \frac{1}{8} u_0 + \frac{3}{4} u_1 + \frac{1}{8} u_2 \right) h - (1) h.
\]

(5.51)
Now divide by \( h \), multiply above and below by 2 in the second term, and rearrange the result to get

\[
\frac{u_2 - u_0}{2h} + \frac{1}{8} (u_0 + 6u_1 + u_2) \frac{2h}{2h} - (1). \tag{5.52}
\]

Therefore, the finite volume method can be interpreted as being composed of a second-order correct finite difference representation of the spacial derivative and a numerical estimate of the function integral \( \int_0^{2h} u dx \).

Figure 5.4: Diagrammatic sketch for representing finite volume formula as a finite difference formula.

### 5.2 Galerkin Method for First Order Equations

In the Galerkin finite element method, we simply change the weighting function \( w_i(x) \) to be the same as the interpolation or basis function. In other words, we let the weighting function be a Lagrange polynomial. To illustrate the concept consider the equation

\[
\frac{du}{dx} - f(x) = 0 \quad x \in [x_0 = a, x_n = b] \tag{5.53}
\]

and the boundary condition

\[
u(0) = u_0. \tag{5.54}\]

From the method of weighted residuals we have

\[
\int R(x) w_i(x) dx = 0 \quad i = 0, \ldots, n. \tag{5.55}
\]

where
\[ R(x) = \frac{d\hat{u}(x)}{dx} - f(x). \]  

(5.56)

Let the weighting function be defined as

\[ w_i(x) = \ell_i(x) \]  

(5.57)

and let us approximate \( u(x) \) as

\[ u(x) \approx \hat{u}(x) = \sum_{j=1}^{n} u_j \ell_j(x). \]  

(5.58)

Then we can substitute Eq. (5.58) into Eq. (5.55) to obtain

\[
\int R(x) w_i(x) \, dx = \int \left[ \frac{d\hat{u}(x)}{dx} - f(x) \right] w_i(x) \, dx \\
= \int \left[ \frac{d\hat{u}(x)}{dx} - f(x) \right] \ell_i(x) \, dx \\
= \int \left[ \sum_{j=0}^{n} \frac{d\hat{u}(x)}{dx} \ell_j(x) - f(x) \right] \ell_i(x) \, dx \\
= \sum_{j=0}^{n} u_j \int \left[ \frac{d\ell_j(x)}{dx} \ell_i(x) \right] \, dx - \int f(x) \ell_i(x) \, dx \\
= 0 \quad i = 0, \ldots, n.
\]  

(5.59)

Let us assume that we will use linear Lagrange polynomial basis functions and three elements. Then \( n = 3 \) (four nodes) and we can write Eq. (5.59) in matrix form as

\[
\begin{bmatrix}
\int \frac{d\ell_0(x)}{dx} \ell_0(x) \, dx & \int \frac{d\ell_0(x)}{dx} \ell_1(x) \, dx & \int \frac{d\ell_0(x)}{dx} \ell_2(x) \, dx & \int \frac{d\ell_0(x)}{dx} \ell_3(x) \, dx \\
\int \frac{d\ell_1(x)}{dx} \ell_0(x) \, dx & \int \frac{d\ell_1(x)}{dx} \ell_1(x) \, dx & \int \frac{d\ell_1(x)}{dx} \ell_2(x) \, dx & \int \frac{d\ell_1(x)}{dx} \ell_3(x) \, dx \\
\int \frac{d\ell_2(x)}{dx} \ell_0(x) \, dx & \int \frac{d\ell_2(x)}{dx} \ell_1(x) \, dx & \int \frac{d\ell_2(x)}{dx} \ell_2(x) \, dx & \int \frac{d\ell_2(x)}{dx} \ell_3(x) \, dx \\
\int \frac{d\ell_3(x)}{dx} \ell_0(x) \, dx & \int \frac{d\ell_3(x)}{dx} \ell_1(x) \, dx & \int \frac{d\ell_3(x)}{dx} \ell_2(x) \, dx & \int \frac{d\ell_3(x)}{dx} \ell_3(x) \, dx
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1 \\
u_2 \\
u_3
\end{bmatrix} = 0
\]  

(5.60)

Consider next the span of the weighting functions. The span is the number of elements over which a Lagrange polynomial is defined. When linear Lagrange polynomials are used as weighting functions the Lagrange is usually defined over two elements, one on each side of the node. This is not true for the first or last nodes (at \( x_0 \) and \( x_n \)) which span only one element. The resulting equation is found below as Eq. (5.61)). Notice that in the second row, for example, there is a zero in the fourth column. This is because the weighting function \( \ell_1(x) \) does not reach the fourth node, that is the one at \( x_3 \). In fact, the maximum number of non-zero coefficients, we will see in a row in a problem of this kind is three and they form a band around the diagonal.
Now we can focus on the integrals that remain. For weighting functions that span at most two elements, we break the integrations into two one-element integrations. Examination of Fig. 5.5 will facilitate understanding the step that takes us from Eq. (5.61) to Eq. (5.62). For completeness, we also express this matrix equation using Lagrange polynomials defined in terms of $\chi$. We obtain

\[
\begin{bmatrix}
\int \frac{d\ell_0(x)}{dx} \ell_0(x) \, dx & \int \frac{d\ell_1(x)}{dx} \ell_0(x) \, dx & 0 & 0 \\
\int \frac{d\ell_0(x)}{dx} \ell_1(x) \, dx & \int \frac{d\ell_1(x)}{dx} \ell_1(x) \, dx & \int \frac{d\ell_2(x)}{dx} \ell_1(x) \, dx & 0 \\
0 & \int \frac{d\ell_0(x)}{dx} \ell_2(x) \, dx & \int \frac{d\ell_1(x)}{dx} \ell_2(x) \, dx & \int \frac{d\ell_3(x)}{dx} \ell_2(x) \, dx \\
0 & 0 & \int \frac{d\ell_1(x)}{dx} \ell_3(x) \, dx & \int \frac{d\ell_2(x)}{dx} \ell_3(x) \, dx \\
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix}
\begin{bmatrix}
\int f(x) \ell_0(x) \, dx \\
\int f(x) \ell_1(x) \, dx \\
\int f(x) \ell_2(x) \, dx \\
\int f(x) \ell_3(x) \, dx
\end{bmatrix}.
\] (5.61)
\[
\begin{bmatrix}
\int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_0(\chi) \frac{d\chi}{dx} \left|_{\chi=1}\right.
& \int_0^1 \frac{d\ell_1(\chi)}{d\chi} \ell_0(\chi) \frac{d\chi}{dx} \left|_{\chi=1}\right. \\
\int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_1(\chi) \frac{d\chi}{dx} \left|_{\chi=1}\right.
& + \int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_0(\chi) \frac{d\chi}{dx} \left|_{\chi=1}\right. \\
0
& \int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_1(\chi) \frac{d\chi}{dx} \left|_{\chi=2}\right. \\
0
& \int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_1(\chi) \frac{d\chi}{dx} \left|_{\chi=3}\right. \\
\end{bmatrix}
\times
\begin{bmatrix}
\int_0^1 f(\chi) \ell_0(\chi) \frac{d\chi}{dx} \left|_{\chi=1}\right. \\
\int_0^1 f(\chi) \ell_1(\chi) \frac{d\chi}{dx} \left|_{\chi=2}\right. \\
\int_0^1 f(\chi) \ell_0(\chi) \frac{d\chi}{dx} \left|_{\chi=2}\right. \\
\int_0^1 f(\chi) \ell_1(\chi) \frac{d\chi}{dx} \left|_{\chi=3}\right. \\
\end{bmatrix}
\]

(5.62)
Since integration is only over elements, for example spacial increments of \([x_1, x_2]\) or \([\chi_0, \chi_1]\), the Lagrange polynomials are all of the same form for this problem

\[
\ell_j(\chi) = \begin{cases} 
\ell_1(\chi) & \chi \in [0, 1] \\
\ell_0(\chi) & \chi \in [0, 1]
\end{cases}.
\] (5.63)

Figure 5.5: Global and local basis functions (Lagrange polynomials) for four-node problem.

We now generate the element coefficient matrix for element 1 that contributes to the matrix on the left-hand side of Eq. (5.62)

\[
\begin{bmatrix}
\int_{x_0}^{x_1} \frac{d\ell_0(x)}{dx} \ell_0(x) dx & \int_{x_0}^{x_1} \frac{d\ell_1(x)}{dx} \ell_0(x) dx \\
\int_{x_0}^{x_1} \frac{d\ell_0(x)}{dx} \ell_1(x) dx & \int_{x_0}^{x_1} \frac{d\ell_1(x)}{dx} \ell_1(x) dx
\end{bmatrix}
\] (5.64)

or using local coordinates

\[
\begin{bmatrix}
\int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_0(\chi) \frac{d\chi}{dx} d\chi & \int_0^1 \frac{d\ell_1(\chi)}{d\chi} \ell_0(\chi) \frac{d\chi}{dx} d\chi \\
\int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_1(\chi) \frac{d\chi}{dx} d\chi & \int_0^1 \frac{d\ell_1(\chi)}{d\chi} \ell_1(\chi) \frac{d\chi}{dx} d\chi
\end{bmatrix}.
\] (5.65)

For element 2, we get

\[
\begin{bmatrix}
\int_{x_0}^{x_2} \frac{d\ell_1(x)}{dx} \ell_1(x) dx & \int_{x_0}^{x_2} \frac{d\ell_2(x)}{dx} \ell_1(x) dx \\
\int_{x_0}^{x_2} \frac{d\ell_1(x)}{dx} \ell_2(x) dx & \int_{x_0}^{x_2} \frac{d\ell_2(x)}{dx} \ell_2(x) dx
\end{bmatrix}
\] (5.66)

or

\[
\begin{bmatrix}
\int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_0(\chi) \frac{d\chi}{dx} d\chi & \int_0^1 \frac{d\ell_1(\chi)}{d\chi} \ell_0(\chi) \frac{d\chi}{dx} d\chi \\
\int_0^1 \frac{d\ell_0(\chi)}{d\chi} \ell_1(\chi) \frac{d\chi}{dx} d\chi & \int_0^1 \frac{d\ell_1(\chi)}{d\chi} \ell_1(\chi) \frac{d\chi}{dx} d\chi
\end{bmatrix}.
\] (5.67)

As noted the element matrices are the same for elements 1, 2, and 3. This would not be the case if the nodes were not equally spaced since the value of \(h\) would change from element to element and this would change the value of the integrals.
As an example of the integration protocol, consider the evaluation of the following two terms that appear in element matrix 1. For the first, we have

\[
\int_{x_0}^{x_1} \frac{d\ell_0(x)}{dx} \ell_0(x) \, dx = \int_0^h \left( -\frac{1}{h} \right) \left( \frac{h-x}{h} \right) \, dx = \left( -\frac{x}{h} + \frac{x^2}{2h^2} \right)|_0^h = -1 + \frac{1}{2} = -\frac{1}{2}
\]

or in local coordinates

\[
\int_0^1 \frac{d\ell_0(\chi)}{d\chi} \frac{d\chi}{dx} \ell_0(\chi) \frac{dx}{d\chi} d\chi = \int_0^1 \left( -\frac{1}{1} \right) \left( \frac{1-x}{1} \right) \left( \frac{h}{1} \right) \, d\chi = \left( -\frac{x}{1} + \frac{x^2}{2} \right)|_0^1 = -\frac{1}{2}
\]

and for the second, we obtain

\[
\int_{x_0}^{x_1} \frac{d\ell_1(x)}{dx} \ell_0 \, dx = \int_0^h \left( \frac{1}{h} \right) \left( \frac{h-x}{h} \right) \, dx = \left( \frac{x}{h} - \frac{x^2}{2h^2} \right)|_0^h = 1 - \frac{1}{2} = \frac{1}{2}
\]
Expressing the above in local coordinates we can write

\[
\int_0^1 d\xi_1(\chi) \frac{d\xi_0(\chi)}{d\chi} \frac{dx}{d\xi} d\chi
\]  

\[
= \int_0^1 \left( \frac{1}{1} \right) \left( \frac{1}{1} \right) \left( \frac{1}{1} \right) \left( \frac{h}{1} \right) d\chi
\]  

\[
= \int_0^1 \left( \frac{1}{1} \right) \left( \frac{1}{1} \right) \left( \frac{h}{1} \right) d\chi
\]  

\[
= \left( \chi - \chi^2 \right) \bigg|_0^1
\]  

\[
= \frac{1}{2}
\]  

(5.71)

Performing similar integrations for the other elements of the element matrix we obtain

\[
\begin{bmatrix}
-\frac{1}{2} & \frac{1}{2} \\
-\frac{1}{2} & \frac{1}{2}
\end{bmatrix}
\]  

(5.72)

and assuming \( f(x) = \alpha \), we get for the term on the right-hand side of Eq. (5.62) (because we assumed \( \alpha \) to be constant over an element)

\[
= \left[ \begin{array}{c} 
\alpha \int_0^1 \ell_1(\chi) \frac{dx}{d\xi} d\chi \\
\alpha \int_0^1 f(\chi) \ell_2(\chi) \frac{dx}{d\xi} d\chi + \alpha \int_0^1 f(\chi) \ell_0(\chi) \frac{dx}{d\xi} d\chi \\
\alpha \int_0^1 f(\chi) \ell_0(\chi) \frac{dx}{d\xi} d\chi + \alpha \int_0^1 f(\chi) \ell_1(\chi) \frac{dx}{d\xi} d\chi \\
\alpha \int_1^0 f(\chi) \ell_1(\chi) \frac{dx}{d\xi} d\chi
\end{array} \right]
\]

\[
= \left[ \begin{array}{c} 
\alpha \int_0^1 \chi h d\chi \\
\alpha \int_0^1 \chi h d\chi + \alpha \int_0^1 \chi h d\chi \\
\alpha \left( \chi - \frac{x^2}{2} \right) h|_0^1 \\
\alpha \left( \chi - \frac{x^2}{2} \right) h|_0^1 + \alpha \left( \chi - \frac{x^2}{2} \right) h|_0^1 \\
\alpha \left( \chi - \frac{x^2}{2} \right) h|_0^1 + \alpha \left( \chi - \frac{x^2}{2} \right) h|_0^1
\end{array} \right]
\]

\[
= \alpha \left[ \begin{array}{c} 
\frac{h}{2} \\
h \\
h \\
h \\
\frac{h}{2}
\end{array} \right]
\]  

(5.73)

If we now replace the integrals in Eq. (5.62) we obtain
5.2. GALERKIN METHOD FOR FIRST ORDER EQUATIONS

\[ \begin{bmatrix}
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2} \\
0 & -\frac{1}{2} \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
0 & 0 \\
\frac{1}{2} & -\frac{1}{2} \\
\frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2}
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1 \\
u_2 \\
u_3
\end{bmatrix}
= \begin{bmatrix}
h\alpha \\
h\alpha \\
h\alpha \\
h\alpha
\end{bmatrix} \] 

(5.74)

or

\[ \begin{bmatrix}
\frac{1}{2} & 0 & 0 \\
\frac{1}{2} & 0 & 0 \\
0 & \frac{1}{2} & 0 \\
0 & 0 & \frac{1}{2}
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1 \\
u_2 \\
u_3
\end{bmatrix}
= \begin{bmatrix}
h\alpha \\
h\alpha \\
h\alpha \\
h\alpha
\end{bmatrix} \] 

(5.75)

The manner in which the element matrices are organized to produce the global coefficient matrix in Eq. (5.74) is illustrated in Fig. 5.6. Each element matrix is indicated by the dashed perimeter.

Figure 5.6: Arrangement of element coefficient matrices to form a global matrix. Each box with a dashed perimeter is an element coefficient matrix.

Another way to look at this information is shown in Fig. 5.7. In this rendition, one can see the interconnectedness of the information from element to element.

Figure 5.7: Selection of matrix elements to illustrate overlap of element information.

To permit a comparison with the analytic solution let \( u_0 = 1, \alpha = 2, \) and \( h = 3; \) then we have
\[
\begin{bmatrix}
0 & 1/2 & 0 \\
-1/2 & 0 & 1/2 \\
0 & -1/2 & 1/2 \\
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix}
= \begin{bmatrix}
(3) (2) + \frac{1}{2} \\
(3) (2) \\
\frac{(3)(2)}{2} \\
\end{bmatrix} 
\] 
(5.76)

or
\[
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix}
= \begin{bmatrix}
7 \\
13 \\
19 \\
\end{bmatrix} . 
\] 
(5.77)

If we now enter these coefficient values into the approximating series for \( \hat{u}(x) \) the result is
\[
\hat{u}(x) = \ell_0(x) + 7\ell_1(x) + 13\ell_2(x) + 19\ell_3(x) . 
\] 
(5.78)

The analytic solution is
\[
u(x) = 2x + 1 .
\] 
(5.79)

If you select an \( x \) value at a node such that only one term in Eq. (5.78) is non-zero it is easy to see that the solution is exact. This is to be expected since the solution is a linear function of \( x \) so it can be exactly represented by a sum of linear Lagrange polynomials.

### 5.2.1 Finite-Difference Interpretation of the Galerkin Approximation

If one expands the matrix product appearing on the left-hand-side of Eq. (5.74) it is seen that the result is a series of finite difference approximations to the first derivative. In viewing the following, keep in mind that each equation is identified with the node that corresponds to the equation. For example, the first equation is identified with an approximation centered at node zero, the second at node 1 etc.

\[
\frac{u_1 - u_0}{h} = \alpha \quad \text{a forward difference} 
\] 
(5.80)

\[
\frac{u_2 - u_0}{2h} = \alpha \quad \text{a centered difference} 
\] 
(5.81)

\[
\frac{u_3 - u_1}{2h} = \alpha \quad \text{a centered difference} 
\] 
(5.82)

\[
\frac{u_3 - u_2}{h} = \alpha \quad \text{a backward difference}. 
\] 
(5.83)

From the above we observe that, but for the first and last nodes in the sequence, the finite element approximation in this particular case is equivalent to a second-order correct finite difference scheme.

### 5.3 Galerkin Method for Second-Order Equations

Consider the second-order equation
\[
\frac{d^2 u(x)}{dx^2} + g_1 \frac{du(x)}{dx} + g_2 u(x) = 0 \quad x \in [x_0, x_n] 
\] 
(5.84)

\[
u(0) = u_0 
\] 
(5.85)

\[
\left. \frac{du(x)}{dx} \right|_{x_n} = u'. 
\] 
(5.86)
As earlier, let the approximation for \( u(x) \) be defined as
\[
  u(x) \approx \hat{u}(x) = \sum_{j=0}^{n} u_j \ell_j(x).
\] (5.87)

Let us define the residual associated with Eq. (5.84) as
\[
  \frac{d^2 \hat{u}(x)}{dx^2} + g_1 \frac{d \hat{u}(x)}{dx} + g_2 \hat{u}(x) = R(x).
\] (5.88)

As in the case of first-order equations, from the theory underlying the Galerkin finite element method we use the basis function as the weighting function to define the residual. We obtain
\[
  \int_{x_0}^{x_n} R(x) \ell_i(x) \, dx = 0 \quad i = 0, 1, \ldots, n.
\] (5.89)

Substitute Eq. (5.88) into (5.89) to obtain
\[
  \int_{x_0}^{x_n} \left( \frac{d^2 \hat{u}(x)}{dx^2} + g_1 \frac{d \hat{u}(x)}{dx} + g_2 \hat{u}(x) \right) \ell_i(x) \, dx = 0 \quad i = 0, 1, \ldots, n.
\] (5.90)

Consider the second-order derivative term in this expression. There is a problem here because, if one looks at the behavior of the second derivative of a linear Lagrange polynomial, one observes that at a node the derivative is discontinuous. While such a discontinuity can be accommodated, it is not convenient to do so.

To avoid this problem we proceed as follows. Using integration by parts for the interval \( x_0 \) to \( x_n \) we have
\[
  \int_{x_0}^{x_n} \frac{d}{dx} \left( \frac{d \hat{u}(x)}{dx} \ell_i(x) \right) \, dx = \int_{x_0}^{x_n} \left( \frac{d \hat{u}(x)}{dx} \frac{d \ell_i(x)}{dx} + \frac{d^2 \hat{u}(x)}{dx^2} \ell_i(x) \right) \, dx \quad i = 0, \ldots, n
\] (5.91)
but
\[
  \int_{x_0}^{x_n} \frac{d}{dx} \left( \frac{d \hat{u}(x)}{dx} \ell_i(x) \right) \, dx = \left. \left( \frac{d \hat{u}(x)}{dx} \ell_i(x) \right) \right|_{x_0}^{x_n}.
\] (5.92)

Now let us consider the right-hand side term in Eq. (5.92). First we expand it to give
\[
  - \left. \frac{d \hat{u}(x)}{dx} \ell_0(x) \right|_{x_0} \quad i = 0
\] (5.93)
and
\[
  \left. \frac{d \hat{u}(x)}{dx} \ell_i(x) \right|_{x_0}^{x_n} \quad i = 0, \ldots, n
\] (5.94)
\[
  \left. \frac{d \hat{u}(x)}{dx} \ell_n(x) \right|_{x_0}^{x_n} \quad i = n.
\] (5.95)

Notice that while there are \( n + 1 \) possible terms in Eq. (5.94), only two survive Eq. (5.95). This is because the values of \( \ell_i(x) \) for \( i \neq 0 \) or \( n \) vanish at \( x = x_0 \) and \( x = x_n \), the endpoints of the interval. Combination of Eq. (5.92) and Eq. (5.91) gives us
\[
  \int_{x_0}^{x_n} \frac{d^2 \hat{u}(x)}{dx^2} \ell_i(x) \, dx = \int_{x_0}^{x_n} \frac{d \hat{u}(x)}{dx} \frac{d \ell_i(x)}{dx} \, dx \quad i = 0, \ldots, n.
\] (5.96)
Let us now use Eq. (5.96) to rewrite Eq. (5.90) as
\[
\int_{x_0}^{x_n} \left( - \frac{d\hat{u}(x)}{dx} \frac{d\ell_i(x)}{dx} + g_1 \frac{d\hat{u}(x)}{dx} \ell_i(x) + g_2 \hat{u}(x) \ell_i(x) \right) dx \\
+ \frac{d\hat{u}(x)}{dx} \ell_n(x) |_{x_n} - \frac{d\hat{u}(x)}{dx} \ell_0(x) |_{x_0} = 0 \quad i = 0, ..., n.
\] (5.97)

Expanding the approximation \( \hat{u}(x) \) and combining the last two terms we obtain
\[
\sum_{j=0}^{n} u_j \int_{x_0}^{x_n} \left( - \frac{d\ell_j(x)}{dx} \frac{d\ell_i(x)}{dx} + g_1 \frac{d\ell_j(x)}{dx} \ell_i(x) + g_2 \ell_j(x) \ell_i(x) \right) dx \\
+ \left( \frac{d\hat{u}}{dx} \ell_i(x) \right) |_{x_0} = 0 \quad i = 0, ..., n.
\] (5.98)

Now consider element-wise integration of the terms in Eq. (5.98) assuming \( g_1 \) and \( g_2 \) to be given constants. For the first element we obtain
\[
\int_{x_0}^{x_1} \left[ - u_0 \frac{d\ell_0(x)}{dx} \frac{d\ell_0(x)}{dx} - u_1 \frac{d\ell_1(x)}{dx} \frac{d\ell_0(x)}{dx} \\
+ g_1 \left( u_0 \frac{d\ell_0(x)}{dx} \ell_0(x) + u_1 \frac{d\ell_1(x)}{dx} \ell_0(x) \right) \\
+ g_2 \left( u_0 \ell_0(x) \ell_0(x) + u_1 \ell_1(x) \ell_0(x) \right) \right] dx
\] (5.99)
or, using local coordinates
\[
\int_{0}^{1} \left[ - u_0 \frac{d\ell_0(\chi)}{d\chi} \frac{d\ell_0(\chi)}{d\chi} - u_1 \frac{d\ell_1(\chi)}{d\chi} \frac{d\ell_0(\chi)}{d\chi} \\
+ g_1 \left( u_0 \frac{d\ell_0(\chi)}{d\chi} \ell_0(\chi) + u_1 \frac{d\ell_1(\chi)}{d\chi} \ell_0(\chi) \right) \\
+ g_2 \left( u_0 \ell_0(\chi) \ell_0(\chi) + u_1 \ell_1(\chi) \ell_0(\chi) \right) \right] \frac{dx}{d\chi} d\chi.
\] (5.100)

If we consider one term in Eq. (5.100) at a time, we can evaluate the integrals defined in the local \( \chi \) coordinate system. In the case of the weighting function \( \ell_0(\chi) \) we obtain for the second-order terms,
\[
- \int_{0}^{1} \left( \frac{d\ell_0(\chi)}{d\chi} \frac{d\ell_0(\chi)}{d\chi} \frac{d\ell_0(\chi)}{d\chi} \right) \frac{dx}{d\chi} d\chi
\] (5.101)
and
\[
- \int_{0}^{1} \left( \frac{d\ell_1(\chi)}{d\chi} \frac{d\ell_0(\chi)}{d\chi} \frac{d\ell_0(\chi)}{d\chi} \right) \frac{dx}{d\chi} d\chi
\] (5.102)
For the first-order terms and using the weighting function $\ell_0(\chi)$ we get

$$
\begin{align*}
\int_0^1 d\ell_0(\chi) \frac{d\chi}{d\chi} & \ell_0(\chi) \frac{dx}{d\chi} d\chi \\
= \int_0^1 \left(-\frac{1}{h}\right) \frac{1}{1} - \chi \frac{h}{1} d\chi \\
= -\chi + \frac{\chi^2}{2} \bigg|_0 = -\frac{1}{2}
\end{align*}
$$

(5.103)

and

$$
\begin{align*}
\int_0^1 d\ell_1(\chi) \frac{d\chi}{d\chi} & \ell_0(\chi) \frac{dx}{d\chi} d\chi \\
= \int_0^1 \left(\frac{1}{h}\right) \frac{1}{1} - \chi \frac{h}{1} d\chi \\
= \chi - \frac{\chi^2}{2} \bigg|_0 = \frac{1}{2}.
\end{align*}
$$

(5.104)

Finally for the zero order terms we obtain

$$
\begin{align*}
\int_0^1 \ell_0(\chi) \ell_0(\chi) \frac{dx}{d\chi} d\chi \\
= \int_0^1 \frac{1}{1} - \chi \frac{h}{1} d\chi \\
= \left(\chi - \chi^2 + \frac{\chi^3}{3}\right) h \bigg|_0 = \frac{h}{3}
\end{align*}
$$

(5.105)

and, finally

$$
\begin{align*}
\int_0^1 \ell_1(\chi) \ell_0(\chi) \frac{dx}{d\chi} d\chi \\
= \int_0^1 \chi \frac{1}{1} - \chi \frac{h}{1} d\chi \\
= \left(\frac{\chi^2}{2} - \frac{\chi^3}{3} h\right) \bigg|_0 = \frac{h}{6}.
\end{align*}
$$

(5.106)

At this point we have evaluations for each term in Eq. (5.100). To see how this fits into the scheme of things, we now generate an element coefficient matrix. Since each element has two nodes and each node is identified with both an equation and the approximation of the unknown value of $\hat{u}(\chi)$, each element coefficient matrix will be $2 \times 2$. In looking at the element coefficient matrices remember that the rows are associated with the weighting functions and the columns are associated with the unknown $u_i$ values. Thus the first rows of the following matrices are associated with the weighting function $\ell_0(\chi)$ and the first columns are associated with the term in the expansion of $\hat{u}(\chi)$ associated
with \( u_0 \):

\[
- \int_0^1 \left[ \frac{d\ell_0(x)}{dx} \frac{dx}{dx} \frac{d\ell_0(x)}{dx} \frac{dx}{dx} \right] \frac{dx}{dx} \frac{dx}{dx} dx + \int_0^1 g_1 \left[ \frac{d\ell_0(x)}{dx} \frac{dx}{dx} \frac{\ell_0(x)}{\ell_1(x)} \frac{dx}{dx} \frac{\ell_0(x)}{\ell_1(x)} \right] \frac{dx}{dx} \frac{dx}{dx} dx + \int_0^1 g_2 \left[ \ell_0(x) \ell_0(x) \ell_1(x) \ell_0(x) \ell_1(x) \right] \frac{dx}{dx} \frac{dx}{dx} dx.
\]  

(5.107)

It is now possible to provide numbers for the coefficients in Eq. (5.107) using Eqs. (5.101) through (5.106). The result is

\[
\left[ \begin{array}{cc}
- \frac{1}{n} & \frac{1}{n}
\end{array} \right] + g_1 \left[ \begin{array}{cc}
- \frac{1}{n} & \frac{1}{n}
\end{array} \right] + g_2 \left[ \begin{array}{cc}
\frac{h}{3} & \frac{h}{3}
\end{array} \right].
\]  

(5.108)

As long as the elements remain the same length and the functions \( g_1 \) and \( g_2 \) remain constant, the above element matrices are all that is needed since each element matrix will be the same. One simply uses the same element matrix for each element along the interval \( x = x_0 \) to \( x = x_n \).

Now consider the last term in Eq. (5.98) for the special case of the equation written for the weighting function \( \ell_0(x) \). Only information in element 1, the first element, has a contribution to this equation, viz.

\[
\left( \frac{d \dot{u}(x)}{dx} \ell_0(x) \right) \bigg|_{x_0}.
\]  

(5.109)

But since \( \ell_0 \) is unity at \( x_0 \)

\[
\left( \frac{d \dot{u}(x)}{dx} \ell_0(x) \right) \bigg|_{x_0} = \frac{d \dot{u}(x)}{dx} \bigg|_{x_0}.
\]  

(5.110)

Since the term \( \left( \frac{d \dot{u}(x)}{dx} \right) \big|_{x_0} \) associated with the flux evaluated at the boundary, it is a flux-type boundary condition. Thus, a second type (also known as a Neumann or flux boundary condition), is imbedded in the equation approximation directly.

Let us now return to our original problem. For interior nodes (not including the boundary nodes of \( 0 \) and \( n \)) we have, using Eqs. (5.90) and (5.91)

\[
\int_{x_0}^{x_n} \left\{ \left( - \frac{d \dot{u}(x)}{dx} \frac{dx}{dx} \frac{d \ell_i(x)}{dx} \right) + \left[ \left( g_1 \frac{d \dot{u}(x)}{dx} + g_2 \frac{d \dot{u}(x)}{dx} \frac{dx}{dx} \right) \frac{dx}{dx} \frac{d \ell_i(x)}{dx} \right] \right\} dx = 0 \quad i = 1, ..., n - 1
\]  

(5.111)

and after substitution of the approximation in Eq. (5.87) into Eq. (5.111) we get

\[
\sum_{j=1}^{n-1} u_j \int_{x_0}^{x_n} \left( \frac{d \ell_j(x)}{dx} \frac{dx}{dx} \frac{d \ell_i(x)}{dx} \right. + g_1 \frac{d \ell_j(x)}{dx} \frac{dx}{dx} \frac{d \ell_i(x)}{dx} + g_2 \frac{d \ell_j(x)}{dx} \frac{dx}{dx} \frac{d \ell_i(x)}{dx} \right) dx = 0 \quad i = 1, ..., n - 1.
\]  

(5.112)

For node zero we have information only from the first element

\[
\sum_{j=0}^{1} u_j \int_{x_0}^{x_1} \left( - \frac{d \ell_j(x)}{dx} \frac{dx}{dx} \frac{d \ell_0(x)}{dx} \right) + \left[ \left( g_1 \frac{d \ell_j(x)}{dx} \frac{dx}{dx} + g_2 \frac{d \ell_j(x)}{dx} \frac{dx}{dx} \right) \frac{dx}{dx} \frac{d \ell_0(x)}{dx} \right] dx - \left( \frac{d \dot{u}(x)}{dx} \ell_0(x) \right) \bigg|_{x_0} = 0.
\]  

(5.113)
Similarly for node $n$ we have

$$
\sum_{j=n-1}^{n} u_j \int_{x_{n-1}}^{x_n} \left( -\frac{d\ell_j (x)}{dx} \frac{d\ell (x)}{dx} \right) + \\
\left[ \left( g_1 \frac{d\ell_j (x)}{dx} + g_2 \ell_j (x) \right) \ell (x) \right] \, dx + \\
\left( \frac{d\hat{u}}{dx} \ell (x)_n \right) |_{x_n} = 0.
$$

(5.115)

This provides us with $n + 1$ equations in $n + 1$ unknowns, which we can assemble into a matrix equation of the form (after accommodation of boundary conditions)

$$
\begin{bmatrix}
  a_{00} & a_{01} & 0 & 0 & 0 & 0 & 0 \\
  a_{10} & a_{11} & a_{12} & 0 & 0 & 0 & 0 \\
  0 & a_{21} & a_{22} & a_{23} & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & a_{n(n-1)} & a_{nn} \\
\end{bmatrix}
\begin{bmatrix}
  u_0 \\
  u_1 \\
  u_2 \\
  u_3 \\
  \vdots \\
  u_n \\
\end{bmatrix}
= 
\begin{bmatrix}
  \frac{d\hat{u}}{dx} \bigg|_{x_0} \\
  0 \\
  0 \\
  0 \\
  0 \\
  0 \\
\end{bmatrix}
$$

(5.116)

where

$$
a_{ij} = \int_{x_0}^{x_n} \left( -\frac{d\ell_j (x)}{dx} \frac{d\ell_i (x)}{dx} + g_1 \frac{d\ell_j (x)}{dx} \ell_i (x) + g_2 \ell_j (x) \right) \, dx
$$

(5.117)

$$
i = 1, \ldots, n-1 \\
u_i = \hat{u} (x_i).
$$

(5.118)

From Eq. (5.86) we have

$$
u (0) = u_0
$$

(5.119)

$$
\left. \frac{d\hat{u} (x)}{dx} \right|_{x_n} = u'
$$

(5.120)

where $u_0$ and $u'$ are given. Substitution of this information into Eq. (5.116), then multiplication of the known value $u_0$ by column 1 and transfer of this information to the right hand side of the equation, and finally elimination of the first row and column, provides

$$
\begin{bmatrix}
  a_{11} & a_{12} & 0 & 0 & 0 & 0 & 0 \\
  a_{21} & a_{22} & a_{23} & 0 & 0 & 0 & 0 \\
  0 & a_{32} & a_{33} & a_{34} & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & 0 & 0 & 0 & a_{n(n-1)} & a_{nn} \\
\end{bmatrix}
\begin{bmatrix}
  u_1 \\
  u_2 \\
  u_3 \\
  \vdots \\
  u_n \\
\end{bmatrix}
= 
\begin{bmatrix}
  -a_{10}u_0 \\
  0 \\
  0 \\
  0 \\
  0 \\
\end{bmatrix}
$$

(5.121)

As shown earlier, the necessary integrations can be performed in the local $\chi$ coordinate system using linear Lagrange functions $\ell_j (\chi)$ $\ell_j = 0, 1$. The integrals for each
term, defined in the global coordinate system are presented below for each of three
cases depending upon the matrix indices. These are the same integrals we evaluated
earlier in Eqs. (5.107) and (5.108) using local coordinates, but now we have combined
information from the two elements associated with a common node to get the equiv-
alent global values. This is evident from the fact that the limits of integration in the
following equations are from \(-h\) to \(h\) and therefore represent a span of two elements.

For the second-order term we have:

\[
\int_{-h}^{h} - \frac{d^2 l_j(x)}{dx} \frac{d^2 l_i(x)}{dx} dx = \frac{1}{h} \quad j < i
\]  
(5.122)

\[
\int_{-h}^{h} - \frac{d^2 l_j(x)}{dx} \frac{d^2 l_i(x)}{dx} dx = -\frac{2}{h} \quad j = i, \quad i, j \neq n, 0
\]  
(5.123)

\[
\int_{-h}^{h} - \frac{d^2 l_j(x)}{dx} \frac{d^2 l_i(x)}{dx} dx = \frac{1}{h} \quad j > i
\]  
(5.124)

\[
\int_{0}^{h} - \frac{d^2 l_j(x)}{dx} \frac{d^2 l_i(x)}{dx} dx = \frac{1}{h} \quad j = i = 0
\]  
(5.125)

\[
\int_{-h}^{0} - \frac{d^2 l_j(x)}{dx} \frac{d^2 l_i(x)}{dx} dx = \frac{1}{h} \quad j = i = n.
\]  
(5.126)

For the first-order term we have:

\[
\int_{-h}^{h} g_1 \frac{d l_j(x)}{dx} l_i(x) dx = -\frac{g_1}{2} \quad j < i
\]  
(5.127)

\[
\int_{-h}^{h} g_1 \frac{d l_j(x)}{dx} l_i(x) dx = 0 \quad j = i, \quad i, j \neq n, 0
\]  
(5.128)

\[
\int_{-h}^{h} g_1 \frac{d l_j(x)}{dx} l_i(x) dx = \frac{g_1}{2} \quad j > i
\]  
(5.129)

\[
\int_{0}^{h} g_1 \frac{d l_j(x)}{dx} l_i(x) dx = -\frac{g_1}{2} \quad j = i = 0
\]  
(5.130)

\[
\int_{-h}^{0} g_1 \frac{d l_j(x)}{dx} l_i(x) dx = \frac{g_1}{2} \quad j = i = n.
\]  
(5.131)

For the zero order term we have:

\[
\int_{-h}^{h} g_2 l_j(x) l_i(x) dx = \frac{g_2}{6} h \quad j < i
\]  
(5.132)

\[
\int_{-h}^{h} g_2 l_j(x) l_i(x) dx = \frac{g_2 h}{3} + \frac{g_2 h}{3} = \frac{2g_2 h}{3} \quad j = i, \quad i, j \neq n, 0
\]  
(5.133)

\[
\int_{-h}^{h} g_2 l_j(x) l_i(x) dx = \frac{g_2 h}{6} \quad j > i
\]  
(5.134)

\[
\int_{0}^{h} g_2 l_j(x) l_i(x) dx = \frac{g_2 h}{3} \quad j = i = 0
\]  
(5.135)

\[
\int_{-h}^{0} g_2 l_j(x) l_i(x) dx = \frac{g_2 h}{3} \quad j = i = n.
\]  
(5.136)

As noted earlier, the coefficients of the known state-variable values (first column of
the matrix) are multiplied by the known state-variable values and the result is placed
on the right-hand side of the equation. The redundant first column and first row in the matrix equation are then removed. Subsequent substitution of the integral values into the resulting element coefficient matrices, and then into the global coefficient matrix, yields the specific form of Eq. (5.121) appropriate for the general form of the problem being considered, that is
\[
\begin{bmatrix}
\frac{1}{h} + 0 + \frac{g_1}{3} & 0 & 0 & \cdots & 0 \\
-\frac{1}{h} + \frac{g_1}{2} + \frac{g_2}{3} & \frac{1}{h} + 0 + \frac{g_2}{6} & 0 & \cdots & 0 \\
0 & -\frac{1}{h} + \frac{g_2}{2} + \frac{g_3}{3} & \frac{1}{h} + 0 + \frac{g_3}{6} & \cdots & 0 \\
0 & 0 & -\frac{1}{h} + \frac{g_3}{2} + \frac{g_4}{3} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \frac{1}{h} + 0 + \frac{g_n}{3} \\
0 & 0 & 0 & \cdots & -\frac{1}{h} + \frac{g_n}{2} + \frac{g_{n+1}}{3}
\end{bmatrix}
\times
\begin{bmatrix}
u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n
\end{bmatrix}
\]

\[
\begin{aligned}
(5.137)
\end{aligned}
\]
To come up with some concrete numbers we assume the following: \( h = 1; \quad g_1 = 1; \quad g_2 = 0.5; \quad u_0 = 1; \quad \frac{du}{dx}\bigg|_{x_n} = 0; \quad n = 3. \) The resulting set of equations is

\[
\begin{bmatrix}
-2 + \frac{1}{3} & 1 + \frac{1}{2} + \frac{1}{12} & 0 \\
1 - \frac{1}{2} + \frac{1}{12} & -2 + \frac{1}{3} & 1 + \frac{1}{2} + \frac{1}{12} \\
0 & 1 - \frac{1}{2} + \frac{1}{12} & -1 + \frac{1}{2} + \frac{1}{6}
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
= \begin{bmatrix}
-1 \left( 1 - \frac{1}{2} + \frac{1}{12} \right) \\
0 \\
0
\end{bmatrix}
\]  

(5.138)

with the solution

\[
[u_1, u_2, u_3]^T = [0.234, -0.121, -0.213]^T.
\]

(5.140)

The analytic solution is

\[
u(x) = \exp(-x/2)[\cos(x/2) + \tan(\pi/4 + 3/2)\sin(x/2)]
\]

(5.141)

and a comparison of numerical and analytical values is presented in the following table:

<table>
<thead>
<tr>
<th>Location</th>
<th>Analytical</th>
<th>Numerical</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x = 1 )</td>
<td>0.200</td>
<td>0.234</td>
<td>0.036</td>
</tr>
<tr>
<td>( x = 2 )</td>
<td>-0.158</td>
<td>-0.121</td>
<td>0.037</td>
</tr>
<tr>
<td>( x = 3 )</td>
<td>-0.241</td>
<td>-0.213</td>
<td>0.028</td>
</tr>
</tbody>
</table>

This is not a bad (but not great) result considering the number of nodes appearing in the approximating equation.

### 5.3.1 Finite Difference Interpretation of Second-Order Galerkin Method

In this subsection we will show how the finite difference and finite element approximations for the second-order equation considered above are related. Let us consider each term in the second row of Eq. (5.137). After each term is divided by \( h \) and the final expression in Eq. (5.144) is multiplied above and below by 2 we obtain

\[
\frac{d^2 u(x)}{dx^2} \approx \frac{1}{h^2} (u_1 - 2u_2 + u_3)
\]

(5.142)

\[
g_1 \frac{du}{dx} \approx -\frac{g_1 u_1}{2h} + (0) u_2 + \frac{g_1 u_3}{2h}
\]

(5.143)

\[
g_2 u(x) \approx \frac{g_2 h}{6h} (u_1 + 4u_2 + u_3) = \frac{g_2}{6} (u_1 + 4u_2 + u_3) \frac{2h}{2h}
\]

(5.144)

Combination of these terms yields

\[
\underbrace{\frac{1}{h^2} (u_1 - 2u_2 + u_3)}_{\text{Centered FD}} + \underbrace{\frac{g_1}{2h} (u_3 - u_1)}_{\text{Centered FD}} + \underbrace{\frac{g_2}{6} (u_1 + 4u_2 + u_3) \frac{2h}{2h}}_{\text{Averaged Simpsons rule for } \int_{x_1}^{x_2} g_2 u(x)dx} = 0
\]

(5.145)

From this equation we see that each full equation (not the ones associated with the end nodes) consists of a second-order correct finite difference approximation to both the first and second-order terms in the original differential equation. For the zero-order term the value of \( \int_{x_1}^{x_2} g_2 u(x)dx \) is obtained using Simpson’s rule divided by the length of the interval over which the integration is being performed, that is \( 2h \).
5.4 Finite Volume Method for Second-Order Equations

We have seen how to formulate the finite volume approximations in one dimension for first-order equations. We now extend that work to consider equations with second derivatives. Consider once again the equation

$$\frac{d^2 u(x)}{dx^2} + g_1 \frac{du(x)}{dx} + g_2 u(x) = 0 \quad x \in [x = x_0, x = x_n]$$

(5.146)

$$u(0) = u_0$$

(5.147)

$$\frac{du}{dx}\bigg|_{x_n} = u'.$$

(5.148)

From the method of weighted residuals we have

$$\int_{x_0}^{x_n} R(x) w_i(x) \, dx = 0 \quad i = 0, 1, ..., n$$

(5.149)

where the weighting function $w_i(x)$ is, as earlier, defined as (see Fig. 5.8)

$$w_i(x) = \begin{cases} 
1 & x \in \left[ x_i - \frac{h}{2}, x_i + \frac{h}{2} \right] \\
0 & x \notin \left[ x_i - \frac{h}{2}, x_i + \frac{h}{2} \right].
\end{cases}$$

(5.150)

The residual is defined as earlier as

$$\begin{align*}
\int_{x_0}^{x_n} \left( \frac{d^2 \hat{u}(x)}{dx^2} + g_1 \frac{d\hat{u}(x)}{dx} + g_2 \hat{u}(x) \right) w_i(x) \, dx = 0 \quad i = 0, 1, ..., n.
\end{align*}$$

(5.152)
Let us now apply integration by parts to the second-order term to obtain

\[
\int_{x_0}^{x_n} \left( \frac{d^2 \hat{u}(x)}{dx^2} + g_1 \frac{d\hat{u}(x)}{dx} + g_2 \hat{u}(x) \right) w_i(x) \, dx
\]

\[
= \int_{x_0}^{x_n} \left\{ \left( - \frac{d\hat{u}(x)}{dx} \frac{dw_i(x)}{dx} \right) + \left[ \left( g_1 \frac{d\hat{u}(x)}{dx} + g_2 \hat{u}(x) \right) w_i(x) \right] \right\} \, dx
\]

\[
+ \frac{d\hat{u}(x)}{dx} w_i|_{x_n} - \frac{d\hat{u}(x)}{dx} w_i|_{x_0}
\]

\[
= 0 \quad i = 0, 1, \ldots, n.
\] (5.153)

Consider the second-order term integrated over the first element, that is

\[- \int_0^h \frac{d\hat{u}(x)}{dx} \frac{dw_i(x)}{dx} \, dx \quad i = 0, 1. \] (5.154)

The term of special interest iE

\[
\frac{dw_i(x)}{dx} \quad i = 0, 1.
\] (5.155)

The derivative of a step function such as \( w_i \) is defined at the point \( x = a \) (as shown in Fig. 5.8) as

\[
\frac{dw_0(x)}{dx} = -\delta (x - a)
\] (5.156)

and

\[
\frac{dw_1(x)}{dx} = +\delta (x - a)
\] (5.157)

where \( \delta_i(x - x_i) \) is the Dirac delta function. The definition of the Dirac delta function is such that

\[
\int_{-\infty}^{\infty} f(x) \delta(x - a) = f(a).
\] (5.158)

Thus

\[- \int_0^h \frac{d\hat{u}(x)}{dx} [-\delta (x - a)] \, dx = \frac{d\hat{u}(x)}{dx} \bigg|_{x=a} \] (5.159)

\[- \int_0^h \frac{d\hat{u}(x)}{dx} [\delta (x - a)] \, dx = -\frac{d\hat{u}(x)}{dx} \bigg|_{x=a}. \] (5.160)

If we now introduce our approximation for \( u(x) \)

\[ u(x) \approx \hat{u}(x) = \sum_{j=0}^{1} u_j \ell_j(x) \] (5.161)

into the right-hand side of Eq. (5.160), the term \( -\frac{d\hat{u}(x)}{dx} \bigg|_{x=a} \) can be written

\[- \frac{d\hat{u}(x)}{dx} \bigg|_{x=a} = - \left[ u_0 \frac{d\ell_0(x)}{dx} + u_1 \frac{d\ell_1(x)}{dx} \right] \bigg|_{x=a}. \] (5.162)

Let us now consider the terms which, in our Galerkin development, gave rise to the imbedded second type boundary conditions. These are, for one element of length \( h \)
which upon expansion gives

\[
\frac{d\hat{u}(x)}{dx} w_0(x) \bigg|_0^h = \frac{d\hat{u}(x)}{dx} \bigg|_h^0 \quad (5.164)
\]

\[
\frac{d\hat{u}(x)}{dx} w_1(x) \bigg|_0^h = \frac{d\hat{u}(x)}{dx} \bigg|_h^1 - \frac{d\hat{u}(x)}{dx} \bigg|_0^0 = \frac{d\hat{u}(x)}{dx} \bigg|_h^1. \quad (5.165)
\]

Note that the terms \( \frac{d\hat{u}(x)}{dx} \bigg|_h^0 \) and \( -\frac{d\hat{u}(x)}{dx} \bigg|_0^0 \) vanish because \( w_0(x) \bigg|_0 \) is 0 and \( w_1(x) \bigg|_0 \) is 0.

We now formulate the coefficient matrices for Eq. (5.153). We assume that the values for each element coefficient matrix are

Element 1

\[
\begin{array}{c|c|c}
\frac{d\hat{u}(x)}{dx} w_0(x) & \frac{d\hat{u}(x)}{dx} w_1(x) \\
\hline
-\int_0^h \frac{d\hat{u}(x)}{dx} w_0(x) \, dx + g_1 \int_0^h \frac{d\hat{u}(x)}{dx} \, dx & -\int_0^h \frac{d\hat{u}(x)}{dx} w_1(x) \, dx + g_1 \int_0^h \frac{d\hat{u}(x)}{dx} \, dx \\
+g_1 \int_0^a \frac{d\hat{u}(x)}{dx} \, dx + g_2 \int_0^a \ell_0(x) \, dx & +g_1 \int_0^a \frac{d\hat{u}(x)}{dx} \, dx + g_2 \int_0^a \ell_1(x) \, dx \\
\end{array}
\]

or in local coordinates

Element 1

\[
\begin{array}{c|c|c}
\frac{d\hat{u}(x)}{dx} w_0(x) & \frac{d\hat{u}(x)}{dx} w_1(x) \\
\hline
\frac{d\hat{u}(x)}{dx} + \epsilon_0(x) & \frac{d\hat{u}(x)}{dx} + \epsilon_1(x) \\
\hline
+g_1 \int_0^a \frac{d\hat{u}(x)}{dx} \, dx & +g_1 \int_0^a \frac{d\hat{u}(x)}{dx} \, dx \\
\end{array}
\]
5.4. FINITE VOLUME METHOD FOR SECOND-ORDER EQUATIONS

Element 2

\[
\begin{align*}
& \begin{array}{c|c|c}
\text{Element 2} & u_1 & u_2 \\
- \int_{h_1}^{2h} \frac{d\ell_1(x)}{dx} \frac{dw_1(x)}{dx} dx + g_1 \int_{b}^{h} \ell_1(x) dx & - \int_{h_1}^{2h} \frac{d\ell_2(x)}{dx} \frac{dw_1(x)}{dx} dx + g_1 \int_{b}^{h} \ell_2(x) dx \\
g_1 \int_{b}^{h} \frac{d\ell_1(x)}{dx} dx + g_2 \int_{b}^{h} \ell_1(x) dx & w_1(x) \\
- \int_{h_1}^{2h} \frac{d\ell_1(x)}{dx} \frac{dw_2(x)}{dx} dx + g_1 \int_{b}^{h} \ell_1(x) dx & - \int_{h_1}^{2h} \frac{d\ell_2(x)}{dx} \frac{dw_2(x)}{dx} dx + g_1 \int_{b}^{h} \ell_2(x) dx \\
g_1 \int_{b}^{h} \frac{d\ell_1(x)}{dx} dx + g_2 \int_{b}^{h} \ell_1(x) dx & w_2(x) \\
\end{array}
\end{align*}
\]

or again using Eqs. (5.154) through (5.160)

\[
\begin{align*}
& \begin{array}{c|c|c}
\text{Element 2} & u_1 & u_2 \\
\frac{d\ell_1(x)}{dx} + g_1 \int_{b}^{h} \frac{d\ell_1(x)}{dx} dx + g_2 \int_{b}^{h} \ell_1(x) dx & \frac{d\ell_2(x)}{dx} + g_1 \int_{b}^{h} \frac{d\ell_2(x)}{dx} dx + g_2 \int_{b}^{h} \ell_2(x) dx \\
\frac{d\ell_1(x)}{dx} + g_1 \int_{b}^{h} \frac{d\ell_1(x)}{dx} dx + g_2 \int_{b}^{h} \ell_1(x) dx & w_1(x) \\
\frac{d\ell_1(x)}{dx} + g_1 \int_{b}^{h} \frac{d\ell_1(x)}{dx} dx + g_2 \int_{b}^{h} \ell_1(x) dx & \frac{d\ell_2(x)}{dx} + g_1 \int_{b}^{h} \frac{d\ell_2(x)}{dx} dx + g_2 \int_{b}^{h} \ell_2(x) dx \\
\frac{d\ell_1(x)}{dx} + g_1 \int_{b}^{h} \frac{d\ell_1(x)}{dx} dx + g_2 \int_{b}^{h} \ell_1(x) dx & w_2(x) \\
\end{array}
\end{align*}
\]

In local coordinates we have

\[
\begin{align*}
& \begin{array}{c|c|c}
\text{Element 2} & u_1 & u_2 \\
\frac{d\ell_0(x)}{dx} \frac{dx}{\alpha x} \int_{0.5}^{\alpha} + g_1 \int_{0.5}^{1} \frac{d\ell_0(x)}{dx} \frac{dx}{\alpha x} d\chi & \frac{d\ell_1(x)}{dx} \frac{dx}{\alpha x} \int_{0.5}^{\alpha} + g_1 \int_{0.5}^{1} \frac{d\ell_1(x)}{dx} \frac{dx}{\alpha x} d\chi \\
+ g_2 \int_{0.5}^{1} \ell_0(x) \frac{dx}{dx} d\chi & + g_2 \int_{0.5}^{1} \ell_1(x) \frac{dx}{dx} d\chi \\
- \frac{d\ell_0(x)}{dx} \frac{dx}{\alpha x} \int_{0.5}^{\alpha} + g_1 \int_{0.5}^{1} \frac{d\ell_0(x)}{dx} \frac{dx}{\alpha x} d\chi & - \frac{d\ell_1(x)}{dx} \frac{dx}{\alpha x} \int_{0.5}^{\alpha} + g_1 \int_{0.5}^{1} \frac{d\ell_1(x)}{dx} \frac{dx}{\alpha x} d\chi \\
+ g_2 \int_{0.5}^{1} \ell_0(x) \frac{dx}{dx} d\chi & + g_2 \int_{0.5}^{1} \ell_1(x) \frac{dx}{dx} d\chi \\
\end{array}
\end{align*}
\]

One observes that, as in earlier examples, given \( h \) is the same in the two elements, the element matrices in local (and also in global) coordinates are identical.

Now we combine the two matrices given by Eqs. (5.167 and (5.170) to create the global matrix. We could also have used the element matrices in local coordinates, that is Eqs. (5.168) and (5.171)
5.4.1 Example of Finite Volume Solution of a Second-Order Equation

Consider the following equation:

$$\frac{d^2 u(x)}{dx^2} + \frac{du(x)}{dx} + 0.5u(x) = 0 \quad x \in [0, 2]$$

$$u(0) = 1$$

$$\left. \frac{du(x)}{dx} \right|_{x=2} = 0.$$  

Using three nodes and the finite volume method we will determine the solution at $x = 0.5$

$$\int_{x_0}^{x_n} R(x) w_i(x) \, dx = 0 \quad i = 0, 1, 2.$$  

The weighting function is, as earlier, defined as (see Fig. 5.9)

$$w_i(x) = \begin{cases} 1 & x \in \left[ x_i - \frac{h}{2}, x_i + \frac{h}{2} \right] \\ 0 & x \notin \left[ x_i - \frac{h}{2}, x_i + \frac{h}{2} \right] \end{cases}.$$
5.4. Finite Volume Method for Second-Order Equations

Figure 5.9: As presented earlier, basis functions \( \ell_j(x) \) and weighting functions \( w_i(x) \) for the finite volume method.

\[
\frac{d^2 \hat{u}(x)}{dx^2} + \frac{d \hat{u}(x)}{dx} + 0.5 \hat{u}(x) = R(x). \tag{5.178}
\]

Substitution of Eq. (5.178) into (5.176) gives

\[
\int_0^2 \left( \frac{d^2 \hat{u}(x)}{dx^2} + \frac{d \hat{u}(x)}{dx} + 0.5 \hat{u}(x) \right) w_i(x) \, dx = 0 \quad i = 0, 1, 2. \tag{5.179}
\]

Let us now use integration by parts to obtain

\[
\int_0^2 \left( \frac{d^2 \hat{u}(x)}{dx^2} + g_1 \frac{d \hat{u}(x)}{dx} + g_2 \hat{u}(x) \right) w_i(x) \, dx
= \int_0^2 \left[ \left( -\frac{d \hat{u}(x)}{dx} \frac{d w_i(x)}{dx} \right) + \left( g_1 \frac{d \hat{u}(x)}{dx} + g_2 \hat{u}(x) \right) w_i(x) \right] \, dx +
\]

\[
\frac{d \hat{u}(x)}{dx} \big|_{x=2} - \frac{d \hat{u}(x)}{dx} \big|_{x=0} = 0 \quad i = 0, 1, 2. \tag{5.180}
\]

Consider the second-order term integrated over the first element, that is

\[
- \int_0^h \frac{d \hat{u}(x)}{dx} \frac{d w_i(x)}{dx} \, dx \quad i = 0, 1. \tag{5.181}
\]

The term of interest is

\[
\frac{d w_i(x)}{dx} \quad i = 0, 1. \tag{5.182}
\]
It is defined as
\[
\frac{dw_0(x)}{dx} = -\delta(x - h/2) \quad (5.183)
\]
and
\[
\frac{dw_1(x)}{dx} = +\delta(x - h/2). \quad (5.184)
\]
Substitution of Eqs. (5.183) and (5.184) into Eq. (5.181) results in
\[
\int_0^h \frac{d\hat{u}(x)}{dx} \left[ -\delta_0(x - h/2) \right] dx = \left. \frac{d\hat{u}(x)}{dx} \right|_{x=h/2}
\]
\[
= \left( u_0 \frac{d\ell_0(x)}{dx} + u_1 \frac{d\ell_1(x)}{dx} \right) \bigg|_{x=h/2} \quad (5.185)
\]
\[
- \int_0^h \frac{d\hat{u}(x)}{dx} \left[ \delta_1(x - h/2) \right] dx = - \left. \frac{d\hat{u}(x)}{dx} \right|_{x=h/2}
\]
\[
= - \left( u_0 \frac{d\ell_0(x)}{dx} + u_1 \frac{d\ell_1(x)}{dx} \right) \bigg|_{x=h/2}. \quad (5.186)
\]
Consider now how to represent the terms
\[
\frac{d\hat{u}(x)}{dx} w_i(x) \bigg|_0^h \quad i = 0, 1. \quad (5.187)
\]
which expanded out, give
\[
\frac{d\hat{u}(x)}{dx} w_0(x) \bigg|_0^h
\]
\[
= \left. \frac{d\hat{u}(x)}{dx} \right|_0 (0) - \left. \frac{d\hat{u}(x)}{dx} \right|_1 (1)
\]
\[
= - \left. \frac{d\hat{u}(x)}{dx} \right|_0 \quad (5.188)
\]
and
\[
\frac{d\hat{u}(x)}{dx} w_1 \bigg|_0^h
\]
\[
= \left. \frac{d\hat{u}(x)}{dx} \right|_0 (1) - \left. \frac{d\hat{u}(x)}{dx} \right|_1 (0)
\]
\[
= \left. \frac{d\hat{u}(x)}{dx} \right|_0. \quad (5.189)
\]
If the term represented in the first line of Eq. (5.188) appears on the left hand side of the domain of interest, it represents a second or Neumann type boundary condition. If a term like that found in Eq. (5.189) is located on the right-hand side of the domain, it also represents a second or Neumann type boundary condition.

Finally, consider now the terms
\[
\int_0^h (0.5) \hat{u}(x) w_0(x) dx = (0.5) \int_0^{h/2} \hat{u}(x) dx = (0.5) \int_0^{h/2} u_0(x) \ell_0 + u_1(x) \ell_1(x) dx
\]
\[
= (0.5) u_0 \int_0^{0.5} \frac{h-x}{h} dx + (0.5) u_1 \int_0^{0.5} \frac{x}{h} dx
\]
\[
= (0.5) (u_0) (3/8) + (0.5) (u_1) (1/8). \quad (5.190)
\]
We now formulate the coefficient matrices for Eq. (5.180) that is.
which in local coordinates becomes

\[
\begin{array}{c|c|c}
\text{Element 1} \\
\hline
\frac{du_0}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_0}{d\chi^2} + \frac{1}{2} \ell_0 (\chi) \right) d\chi & \quad & \frac{du_1}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_1}{d\chi^2} + \frac{1}{2} \ell_1 (\chi) \right) d\chi \quad w_0 (\chi)
\end{array}
\]

\begin{align}
\frac{du_0}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_0}{d\chi^2} + \frac{1}{2} \ell_0 (\chi) \right) d\chi & = w_0 (\chi) \\
\frac{du_1}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_1}{d\chi^2} + \frac{1}{2} \ell_1 (\chi) \right) d\chi & = w_1 (\chi)
\end{align}

Consider also element 2

\[
\begin{array}{c|c|c}
\text{Element 2} \\
\hline
\frac{d^2u_1}{d\chi^2} |_{1.5} + \int_{1.5}^{2.5} \left( \frac{d^2u_1}{d\chi^2} + \frac{1}{2} \ell_1 (\chi) \right) d\chi & \quad & \frac{d^2u_2}{d\chi^2} |_{1.5} + \int_{1.5}^{2.5} \left( \frac{d^2u_2}{d\chi^2} + \frac{1}{2} \ell_2 (\chi) \right) d\chi \quad w_1 (\chi)
\end{array}
\]

\begin{align}
\frac{d^2u_1}{d\chi^2} |_{1.5} + \int_{1.5}^{2.5} \left( \frac{d^2u_1}{d\chi^2} + \frac{1}{2} \ell_1 (\chi) \right) d\chi & = w_1 (\chi) \\
\frac{d^2u_2}{d\chi^2} |_{1.5} + \int_{1.5}^{2.5} \left( \frac{d^2u_2}{d\chi^2} + \frac{1}{2} \ell_2 (\chi) \right) d\chi & = w_2 (\chi)
\end{align}

which using local coordinates yields

\[
\begin{array}{c|c|c}
\text{Element 2} \\
\hline
\frac{du_1}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_1}{d\chi^2} + \frac{1}{2} \ell_1 (\chi) \right) d\chi & \quad & \frac{du_2}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_2}{d\chi^2} + \frac{1}{2} \ell_2 (\chi) \right) d\chi \quad w_0 (\chi)
\end{array}
\]

\begin{align}
\frac{du_1}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_1}{d\chi^2} + \frac{1}{2} \ell_1 (\chi) \right) d\chi & = w_0 (\chi) \\
\frac{du_2}{d\chi} |_{0.5} + \int_{0.5}^{1.5} \left( \frac{d^2u_2}{d\chi^2} + \frac{1}{2} \ell_2 (\chi) \right) d\chi & = w_1 (\chi)
\end{align}
Since the element matrices are identical in local coordinates (because the space increment is constant), we will consider only the element 1 matrix hereinafter.

We first introduce the local Lagrange polynomials and the derivative transforms evaluated for this problem into Eqs. (5.192) and (5.194) to obtain

Element 1

\[
\begin{align*}
&\int_0^{0.5} \left( \frac{1}{2} \chi \right) \left( \frac{1}{2} \right) \, d\chi & \int_0^{0.5} \left( \frac{1}{2} \chi \right) \left( \frac{1}{2} \right) \, d\chi & w_0(\chi) \\
&\int_0^{0.5} \left( \frac{1}{2} \chi \right) \left( \frac{1}{2} \right) \, d\chi & \int_0^{0.5} \left( \frac{1}{2} \chi \right) \left( \frac{1}{2} \right) \, d\chi & w_1(\chi)
\end{align*}
\]

\begin{equation}
(5.195)
\end{equation}

Element 1

\[
\begin{align*}
&-1 + \int_0^{0.5} \left( -1 + \frac{1}{2} (1 - \chi) \right) \, d\chi & 1 + \int_0^{0.5} \frac{1}{2} \left( 1 + \frac{1}{2} (1 - \chi) \right) \, d\chi & w_0(\chi) \\
&1 + \int_0^{0.5} \left( -1 + \frac{1}{2} (1 - \chi) \right) \, d\chi & -1 + \int_0^{0.5} \frac{1}{2} \left( 1 + \frac{1}{2} (1 - \chi) \right) \, d\chi & w_1(\chi)
\end{align*}
\]

\begin{equation}
(5.196)
\end{equation}

Now we combine the two element matrices to create the global matrix

\[
\begin{bmatrix}
-1 \frac{1}{2} + \frac{3}{16} & 1 \frac{1}{2} + \frac{1}{16} & 0 \\
0 & \frac{1}{2} + \frac{1}{16} & 1 \frac{1}{2} + \frac{3}{16} & -1 \frac{1}{2} + \frac{3}{16}
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1
\end{bmatrix}
\]

\begin{equation}
(5.197)
\end{equation}

or, adding up the components of each term we get

\[
\begin{bmatrix}
-21 & 25 & 0 \\
9 & -26 & 25 \\
0 & 9 & -5
\end{bmatrix}
\begin{bmatrix}
u_0 \\
u_1 \\
u_2
\end{bmatrix} = \begin{bmatrix}
-du_0 \\
\frac{dx}{dx}
\end{bmatrix}.
\]

\begin{equation}
(5.198)
\end{equation}

Now we impose the boundary conditions and reduce the matrix to get

\[
\begin{bmatrix}
-26 & 25 \\
9 & -5
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2
\end{bmatrix} + \frac{1}{16} \begin{bmatrix}
9 (1) \\
0
\end{bmatrix} = \begin{bmatrix}
0 \\
0
\end{bmatrix}.
\]

\begin{equation}
(5.199)
\end{equation}

Taking the inverse and rearranging the equation, we arrive at the solution for \( u_1 \) and \( u_2 \), that is
To obtain a solution at \( \hat{u}(0.5) \) you use the approximation for the first element, that is
\[
\hat{u}(0.5) = \ell_0(0.5) u_0 + \ell_1(0.5) u_1
\] (5.204)
or
\[
\hat{u}(0.5) = \frac{1}{2}u_0 + \frac{1}{2}u_1
\]
\[
= (1/2)1 - (1/2)0.47
\]
\[
= 0.
\] (5.205)
The analytical solution is
\[
u(x) = \exp(-x/2)[\cos(x/2) + \tan(\pi/4 + 1)\sin(x/2)].
\] (5.206)
This gives
\[
u(x = 0.5) = -0.129
\] (5.207)
\[
u(x = 1) = -0.802
\] (5.208)
\[
u(x = 2) = -0.122.
\] (5.209)

A comparison between the analytic and numerical results is given in the following table.

<table>
<thead>
<tr>
<th></th>
<th>Analytic</th>
<th>Numerical</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>-0.129</td>
<td>0.265</td>
<td>0.384</td>
</tr>
<tr>
<td>1</td>
<td>-0.802</td>
<td>-0.474</td>
<td>0.328</td>
</tr>
<tr>
<td>2</td>
<td>-1.22</td>
<td>-0.853</td>
<td>0.36</td>
</tr>
</tbody>
</table>

and in Fig. 5.10

This is not a particularly impressive result and this is due to the form of the function \( u(x) \) and the way the third term in the differential equations is approximated numerically. To provide confidence that the algorithm provides a solution that does, indeed, converge to the correct solution as \( h \) approaches zero, we plot in Fig. 5.11 the relationship between the error between the analytical and numerical solutions and \( h \) as \( h \) decreases. It is evident from this graph that the method is second-order accurate (the slope of the line is 2.0) and the error is decreasing as the square of \( h \) as expected. See Section 4.2 on page 69 for a discussion of convergence.
5.4.2 Finite Difference Representation of the Finite-Volume Method for Second-Order Equations

Consider the equation written for the weighting function $w_1$. It is formed by the middle row in Eq. (5.172), that is,

$$
\left[ \frac{1}{h} + \int_{h/2}^{h} \left( -\frac{1}{h} + \frac{h-x}{h} \right) \, dx \right] u_0 \\
+ \left[ \frac{1}{h} + \int_{h/2}^{h} \left( \frac{1}{h} + \frac{x}{h} \right) \, dx + \int_{0}^{h/2} \left( -\frac{1}{h} + \frac{h-x}{h} \right) \, dx \right] u_1 \\
+ \left[ \frac{1}{h} + \int_{0}^{h/2} \left( \frac{1}{h} + \frac{x}{h} \right) \, dx \right] u_2 = 0.
$$

(5.210)

Upon integration we obtain

$$
\left( \frac{1}{h} - \frac{h}{2h} + \frac{h^3}{8h} \right) u_0 + \left( \frac{2}{h} + 0 + \frac{6h}{8} \right) u_1 + \left( \frac{1}{h} + \frac{h}{2h} + \frac{h^2}{8h} \right) u_2 = 0.
$$

(5.211)

Rearranging this expression we get

$$
\frac{1}{h^2} (u_0 - 2u_1 + u_2) + \frac{1}{2h} (u_2 - u_0) + h \left( \frac{1}{8} u_0 + \frac{6}{8} u_1 + \frac{1}{8} u_2 \right) = 0.
$$

If we divide this equation by $h$ we have

$$
\frac{1}{h^2} \text{ centered FD} + \frac{1}{2h} \text{ centered FD} + \left( \frac{1}{8} u_0 + \frac{6}{8} u_1 + \frac{1}{8} u_2 \right) \frac{2h}{2h} = 0.
$$

(5.212)

We see from Eq. (5.212) that the finite volume method applied to this one-dimensional equation is equivalent to a second-order difference approximation for the second and first derivatives and an integrated average over the two element interval for $u(x)$. One feature of the integration formula is that with the $\frac{1}{8}, \frac{3}{4}, \frac{1}{8}$ weighting there is more emphasis put on the diagonal element of the matrix that with the $\frac{1}{6}, \frac{2}{3}, \frac{1}{6}$ weighting of Simpson’s rule encountered in the Galerkin finite element method.
5.5. COLLOCATION METHOD

5.5 Collocation Method

5.5.1 Collocation Method for First-Order Equations

In this section we introduce the collocation method. As will be seen, this is a very attractive method inasmuch as it is not necessary to perform any integrations. The downside of the method is that it requires the use of Hermite polynomials, rather than Lagrange polynomials as basis functions for second-order derivatives. In the case of first-order equations as considered in this subsection, we will not require Hermites. So, starting with first-order equations, let us get started. Consider the equation

\[ \frac{du(x)}{dx} + gu(x) + f(x) = 0 \quad u(0) = u_0 \quad x \in [0, 1]. \]  

(5.213)

As usual, let \( u(x) \) be represented as

\[ u(x) = \hat{u}(x) + E(x) \]  

(5.214)

and

\[ \hat{u}(x) = \sum_{j=0}^{n} u_j \ell_j(x). \]  

(5.215)

Further, let the residual \( R(x) \) be defined as

\[ R(x) \equiv \frac{d\hat{u}(x)}{dx} + g\hat{u}(x) + f(x). \]  

(5.216)

From the method of weighted residuals

\[ \int_{0}^{1} R(x) w_i(x) \, dx = 0 \quad i = 0, 1, \ldots, n. \]  

(5.217)
The bounds on the integral stem from the definition of the domain as presented in Eq. (5.213).

Let us choose the weighting function \( w_i(x) \) to be

\[
w_i(x) = \delta(x - \bar{x}_i)
\]

(5.218)

where \( \delta(x - \bar{x}_i) \) is the Dirac delta function. This choice defines the collocation method.

From the definition of the Dirac delta function, we can write the following:

\[
\int_0^1 R(x) \, w_i(x) \, dx = \int_0^1 R(x) \, \delta(x - \bar{x}_i) \, dx = R(\bar{x}_i).
\]

(5.219)

Then our equations for the collocation method are obtained from

\[
R(\bar{x}_i) = 0 \quad i = 0, 1, ..., n
\]

(5.220)

where \( \bar{x}_i \) are called collocation points. Optimal accuracy is achieved by letting the collocation points be the Gauss points. These are the same points introduced in our discussion of Gaussian Legendre quadrature integration in Section 3.1.6 on page 59. The required point values of these points are tabulated in reference books.

Next substitute for \( R(x) \) using Eq. (5.216) to give

\[
\int_0^1 \frac{d\ell_j(x)}{dx} + g \ell_j(x) + f(x) \delta(x - \bar{x}_i) \, dx = 0 \quad i = 0, 1, ..., n.
\]

(5.221)

Now substitute the definition of \( \hat{u}(x) \) using Eq. (5.215) to give

\[
\sum_{j=0}^n u_j \left[ \frac{d\ell_j(x)}{dx} + g \ell_j(x) \right]_{x=\bar{x}_i} + f(\bar{x}_i) = 0 \quad i = 0, 1, ..., n.
\]

(5.222)

We can now write Eq. (5.222) using matrix notation as

\[
\begin{bmatrix}
a_{11} & a_{12} & 0 & 0 & 0 & 0 \\
a_{21} & a_{22} & a_{23} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & a_{n(n-1)} & a_{nn} & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
\vdots \\
u_n
\end{bmatrix}
= \begin{bmatrix}
-e_1 - a_{10} u_0 \\
-e_2 \\
\vdots \\
-e_n
\end{bmatrix}
\]

(5.223)

where

\[
a_{ij} = \left[ \frac{d\ell_j(x)}{dx} + g \ell_j(x) \right]_{x=\bar{x}_i}
\]

(5.224)

and

\[
e_i = f(\bar{x}_i).
\]

(5.225)
5.5. COLLOCATION METHOD

After substitution of the values of $\ell_j(x)$, Eq. (5.223)

$$\hat{u}(x) = \sum_{j=0}^{2} u_j \ell_j(x). \quad (5.226)$$

Now evaluate $R(\bar{x}_i) \quad i = 0, 1$ at the Gauss points and set the result to zero for each point. We obtain

$$R(\bar{x}_i) = 0 \quad i = 0, 1 \quad (5.227)$$

or

$$\left[ \frac{d\hat{u}(x)}{dx} + \frac{0.9}{1+2x} \hat{u}(x) \right]_{\bar{x}_i} = 0 \quad i = 0, 1. \quad (5.228)$$

Next substitute for $\hat{u}(x)$ using Eq. (5.226) to get

$$\sum_{j=0}^{2} u_j \left[ \frac{d\ell_j(x)}{dx} + \frac{0.9}{1+2x} \ell_j(x) \right]_{\bar{x}_i} = 0 \quad i = 0, 1. \quad (5.229)$$

Figure 5.12: Element array for collocation method explanation.

Now, using the Gauss Legendre integration points selected from Table 3.29. Because our basis functions are piecewise linear we require one Gaussian point per element and these points are located at $\bar{x}_0 = 0.25$ and $\bar{x}_1 = 0.75$ (keep in mind that the table assumes a $[-1,1]$ coordinate system for each element). From Eq. (5.229) we obtain the following matrix equation:

$$\begin{bmatrix}
\frac{d\ell_0(x)}{dx} + \frac{0.9}{1+2x} \ell_0(x) \\
\frac{d\ell_1(x)}{dx} + \frac{0.9}{1+2x} \ell_1(x) \\
\frac{d\ell_2(x)}{dx} + \frac{0.9}{1+2x} \ell_2(x)
\end{bmatrix}_{\bar{x}=0.25} \times 
\begin{bmatrix}
u_0 \\
u_1 \\
u_2
\end{bmatrix} = 
\begin{bmatrix}0 \\
0 \\
0
\end{bmatrix}. \quad (5.230)$$
Notice that this is a rectangular matrix, a phenomenon identified with collocation. No integrations appear in Eq. (5.230), so the coefficients are easily evaluated by inspection of the function values via Fig. 5.12. We obtain through substitution of the collocation point locations

\[
\begin{pmatrix}
\frac{1}{0.9} + 0.5 \\
\frac{-1}{0.9} + 0.5 \\
0 \\
\end{pmatrix}
\begin{pmatrix}
\frac{1}{0.9} + 0.5 \\
\frac{-1}{0.9} + 0.5 \\
0 \\
\end{pmatrix}
\begin{pmatrix}
0 \\
0 \\
0 \\
\end{pmatrix}
\begin{pmatrix}
u_0 \\
u_1 \\
u_2 \\
\end{pmatrix}
= \begin{pmatrix} 0 \\
0 \\
0 \\
\end{pmatrix}
\]  
(5.231)

or

\[
\begin{pmatrix}
-1.7 & 2.3 & 0 \\
0 & -1.82 & 2.18 \\
\end{pmatrix}
\begin{pmatrix}
u_0 \\
u_1 \\
u_2 \\
\end{pmatrix}
= \begin{pmatrix} 0 \\
0 \\
0 \\
\end{pmatrix}.
\]  
(5.232)

The rectangular form of the coefficient matrix would normally be problematic, but we have a solution to the problem. We now impose the boundary condition which states that \( u_0 = 1 \). After multiplying the first column by this value and transferring the information to the right-hand side we have

\[
\begin{pmatrix}
2.3 & 0 \\
-1.82 & 2.18 \\
\end{pmatrix}
\begin{pmatrix}
u_1 \\
u_2 \\
\end{pmatrix}
= \begin{pmatrix} -(1.7) (1) \\
0 \\
\end{pmatrix} 
\]  
(5.233)

which is a solvable system of equations. From this equation we get

\[
\begin{pmatrix}
u_1 \\
u_2 \\
\end{pmatrix}
= \begin{pmatrix} 0.740 \\
0.618 \\
\end{pmatrix} 
\]  
(5.234)

while the analytical solution is

\[
\begin{pmatrix}
u_1 \\
u_2 \\
\end{pmatrix}
= \begin{pmatrix} 0.732 \\
0.609 \\
\end{pmatrix} 
\]  
(5.235)

which is quite accurate considering we are using only three nodes.

### 5.5.2 Collocation Method for Second-Order Equations

In this subsection we will extend our discussion of the collocation method to consider second derivatives. It is at this point that we will need to introduce the Hermite polynomials as basis functions. The Hermite polynomial was introduced earlier in Subsection 1.9 on page 20.

Consider now the second-order equation

\[
\frac{d^2u(x)}{dx^2} + \frac{du(x)}{dx} + 0.5u(x) = 0.
\]  
(5.236)

Following the strategy we used in the preceding subsection, we define our weighted residual operator as

\[
\int_0^1 R(x) w_i(x) dx = \int_0^1 R(x) \delta(x - \bar{x}_i) dx = R(\bar{x}_i) = 0 \quad i = 0, ..., n
\]  
(5.237)

where the residual is defined as

\[
R(x) = \frac{d^2u(x)}{dx^2} + \frac{du(x)}{dx} + 0.5u(x)
\]  
(5.238)
from which we get

\[
\left[ \frac{d^2 \hat{u}(x)}{dx^2} + \frac{d \hat{u}(x)}{dx} + 0.5 \hat{u}(x) \right]_{\hat{x}_i} = 0 \quad i = 0, 1, \ldots, n. \tag{5.239}
\]

We see immediately that if we use linear functions to represent \( u(x) \) we will have difficulty because the second derivative is zero at the collocation points. Your first inclination probably is to use quadratic Lagrange polynomials, since their second derivatives will be defined everywhere except at the end-of-the-element nodes. However, one cannot ignore these end points since the collocation method does not have restrictions on where the collocation points should be. As noted earlier, the simplest answer is to use Hermite polynomials which we introduced in Section 1.9 on page 20, that is use

\[
\hat{u}(x) = \sum_{j=0}^{n} h^0_j(x) u_j + h^1_j(x) \frac{du_j}{dx} \tag{5.240}
\]

(see Fig. 5.13) which provides us with \( 2(n+1) \) unknowns. The Hermite basis functions are defined as

\[
h^0_j(x) = (\ell_j(x))^2 \left( 1 - 2(x-x_j) \frac{d\ell_j}{dx} \right) \tag{5.241}
\]

and

\[
h^1_j(x) = (\ell_j(x))^2 (x-x_j) \tag{5.242}
\]

where, for convenience, the Lagrange polynomial approximations are used in the definition.

To make the notation tolerable, let us define the following operator:

\[
L(\cdot) \equiv \frac{d^2 (\cdot)}{dx^2} + \frac{d (\cdot)}{dx} + 0.5 (\cdot). \tag{5.243}
\]

where you can replace the dot with any suitable function, for example \( u(x) \). Using this operator, and assuming two collocation points per element, our collocation equations become

\[
L(\hat{u})_{\hat{x}_i} = 0 \quad i = 0, 1, \ldots, (2n - 1) \tag{5.244}
\]

or

\[
L \left( \sum_{j=0}^{n} h^0_j(x) u_j + h^1_j(x) \frac{du_j}{dx} \right)_{\hat{x}_i} = 0 \quad i = 0, 1, \ldots, (2n - 1) \tag{5.245}
\]

which gives us \( 2n \) equations, which as we will see, is the number we need after implementation of boundary conditions.

If we now write Eq. (5.245) for the specific system of functions and collocation points shown in Figs 5.14 and 5.15 as a matrix equation, we obtain
Figure 5.13: Cubic Hermite polynomials.

\[
\begin{bmatrix}
Lh_0^0 (x) | x_1 & Lh_1^0 (x) | x_1 & Lh_0^1 (x) | x_1 & Lh_1^1 (x) | x_1 & 0 & 0 \\
Lh_0^0 (x) | x_2 & Lh_1^0 (x) | x_2 & Lh_0^1 (x) | x_2 & Lh_1^1 (x) | x_2 & 0 & 0 \\
0 & 0 & Lh_0^0 (x) | x_3 & Lh_1^0 (x) | x_3 & Lh_0^1 (x) | x_3 & Lh_1^1 (x) | x_3 \\
0 & 0 & 0 & Lh_0^0 (x) | x_4 & Lh_1^0 (x) | x_4 & Lh_0^1 (x) | x_4 & Lh_1^1 (x) | x_4 \\
\end{bmatrix}
\times
\begin{bmatrix}
u_0 \\ u_1 \\ \frac{du_0}{dx} \\ \frac{du_1}{dx} \\ u_2 \\ \frac{du_2}{dx} \\
\end{bmatrix}
= \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\
\end{bmatrix}
\] (5.246)

where

\[
L \left[ h_0^0 (x) \right] | \bar{x}_1 \equiv \left[ \frac{d^2 h_0^0 (x)}{d x^2} + \frac{d h_0^0 (x)}{d x} + h_0^0 (x) \right] | \bar{x}_1.
\] (5.247)

Note that the Hermite polynomials are defined element-wise and that the collocation points \( \bar{x}_1 \) and \( \bar{x}_2 \) are in element 1 and \( \bar{x}_3 \) and \( \bar{x}_4 \) are in element 2.

Assume that the boundary conditions on this problem are

\[
u (0) = \bar{u} \hspace{1cm} (5.248)
\]

\[
\frac{du (x)}{dx} \bigg|_{x=2} = \frac{d \bar{u}}{dx} \hspace{1cm} (5.249)
\]

Then Eq. (5.246) becomes, after reducing the set of equations to accommodate these
boundary conditions

\[
\begin{bmatrix}
L (h_0^1 (x)) |_{x_1} & L (h_1^0 (x)) |_{x_1} & L (h_1^1 (x)) |_{x_1} & 0 \\
L (h_0^1 (x)) |_{x_2} & L (h_1^0 (x)) |_{x_2} & L (h_1^1 (x)) |_{x_2} & 0 \\
0 & L (h_0^1 (x)) |_{x_3} & L (h_1^0 (x)) |_{x_3} & L (h_1^1 (x)) |_{x_3} \\
0 & L (h_0^1 (x)) |_{x_4} & L (h_1^0 (x)) |_{x_4} & L (h_1^1 (x)) |_{x_4}
\end{bmatrix}
\begin{bmatrix}
\frac{du_0}{dx} \\
u_1 \\
\frac{du_1}{dx} \\
u_2
\end{bmatrix}
\]

(5.250)

or

\[
[A] \{u\} = \{f\}
\]

(5.251)

from which we get formally

\[
\{u\} = [A]^{-1} \{f\}.
\]

(5.252)

Substitution of \{u\} into Eq. (5.240) gives the required approximate solution across the entire domain:

Let us now apply the collocation method to the equation we used earlier, namely
\[
\frac{d^2 u(x)}{dx^2} + \frac{du(x)}{dx} + \frac{u(x)}{2} = 0 \quad x \in [x = 0, x = 2] \tag{5.253}
\]

\[
u(0) = 1 \tag{5.254}
\]

\[
\frac{du(x)}{dx} \bigg|_{x=2} = 0. \tag{5.255}
\]

To solve this problem we will introduce the coordinate system \(-1 \leq \xi \leq 1\) which is the same coordinate system we used in our discussion of Gauss-Legendre quadrature integration (Section 3.1.6 on page 59, but there we used the \(x\) notation. In this coordinate system, the collocation equations become, for a problem involving only three nodes (two elements)

\[
\begin{bmatrix}
L (h_0^1 (\xi)) |_{x_1(\xi)} & L (h_0^0 (\xi)) |_{x_2(\xi)} & L (h_0^1 (\xi)) |_{x_3(\xi)} & L (h_0^1 (\xi)) |_{x_4(\xi)} & 0
\end{bmatrix}
\begin{bmatrix}
du_0 \\
du_1 \\
du_2 \\
du_3 \\
du_4
\end{bmatrix}
\]

\[
= \begin{bmatrix}
L (h_0^0 (\xi)) |_{x_1(\xi)} & L (h_0^0 (\xi)) |_{x_2(\xi)} & L (h_0^0 (\xi)) |_{x_3(\xi)} & L (h_0^0 (\xi)) |_{x_4(\xi)}
L (h_1^0 (\xi)) |_{x_1(\xi)} & L (h_1^0 (\xi)) |_{x_2(\xi)} & L (h_1^0 (\xi)) |_{x_3(\xi)} & L (h_1^0 (\xi)) |_{x_4(\xi)}
L (h_1^1 (\xi)) |_{x_1(\xi)} & L (h_1^1 (\xi)) |_{x_2(\xi)} & L (h_1^1 (\xi)) |_{x_3(\xi)} & L (h_1^1 (\xi)) |_{x_4(\xi)}
L (h_1^1 (\xi)) |_{x_1(\xi)} & L (h_1^1 (\xi)) |_{x_2(\xi)} & L (h_1^1 (\xi)) |_{x_3(\xi)} & L (h_1^1 (\xi)) |_{x_4(\xi)}
\end{bmatrix}
\]

where

\[
L (\cdot) = \frac{d^2 (\cdot)}{dx^2} + \frac{d (\cdot)}{dx} + \frac{(\cdot)}{2}.
\]

Note that the Hermite polynomials have a very simple form in this \(\xi\) coordinate system, that is

\[
h_0^0 (\xi) = \frac{1}{4} (\xi - 1)^2 (\xi + 2) \tag{5.257}
\]

\[
h_0^1 (\xi) = -\frac{1}{4} (\xi + 1)^2 (\xi - 2) \tag{5.258}
\]

\[
h_1^0 (\xi) = \frac{h}{8} (\xi + 1) (\xi - 1)^2 \tag{5.259}
\]

\[
h_1^1 (\xi) = \frac{h}{8} (\xi + 1)^2 (\xi - 1). \tag{5.260}
\]

To approximate Eq. (5.253) we will need the first and second derivatives of Eq. (5.257) through Eq. (5.260), that is

\[
\frac{dh_0^0 (\xi)}{dx} = \frac{dh_0^0 (\xi)}{d\xi} \frac{d\xi}{dx} = 2 \frac{dh_0^0 (\xi)}{d\xi} \tag{5.261}
\]

\[
\frac{dh_0^1 (\xi)}{dx} = \frac{dh_0^1 (\xi)}{d\xi} \frac{d\xi}{dx} = 2 \frac{dh_0^1 (\xi)}{d\xi} \tag{5.262}
\]

\[
\frac{dh_1^0 (\xi)}{dx} = \frac{dh_1^0 (\xi)}{d\xi} \frac{d\xi}{dx} = 2 \frac{dh_1^0 (\xi)}{d\xi} \tag{5.263}
\]

\[
\frac{dh_1^1 (\xi)}{dx} = \frac{dh_1^1 (\xi)}{d\xi} \frac{d\xi}{dx} = 2 \frac{dh_1^1 (\xi)}{d\xi} \tag{5.264}
\]
5.5. *Collocation Method*

and

\[
\frac{d^2 h_0^0 (\xi)}{d^2 x} = \frac{d^2 h_0^0 (\xi)}{d^2 \xi} \frac{d \xi}{d x} \frac{d \xi}{d x} = 4 \frac{d^2 h_0^0 (\xi)}{d^2 \xi} \tag{5.265}
\]

\[
\frac{d^2 h_0^0 (\xi)}{d^2 x} = \frac{d^2 h_0^0 (\xi)}{d^2 \xi} \frac{d \xi}{d x} \frac{d \xi}{d x} = 4 \frac{d^2 h_0^0 (\xi)}{d^2 \xi} \tag{5.266}
\]

\[
\frac{d^2 h_0^0 (\xi)}{d^2 x} = \frac{d^2 h_0^0 (\xi)}{d^2 \xi} \frac{d \xi}{d x} \frac{d \xi}{d x} = 4 \frac{d^2 h_0^0 (\xi)}{d^2 \xi} \tag{5.267}
\]

\[
\frac{d^2 h_1^1 (\xi)}{d^2 x} = \frac{d^2 h_1^1 (\xi)}{d^2 \xi} \frac{d \xi}{d x} \frac{d \xi}{d x} = 4 \frac{d^2 h_1^1 (\xi)}{d^2 \xi} \tag{5.268}
\]

where the first derivatives are, from Eq. (5.257) through (5.260)

\[
\frac{dh_0^0 (\xi)}{d \xi} = \frac{1}{4} \left[ (\xi - 1)^2 + 2 (\xi + 2) (\xi - 1) \right] \tag{5.269}
\]

\[
\frac{dh_0^1 (\xi)}{d \xi} = -\frac{1}{4} \left[ (\xi + 1)^2 + 2 (\xi - 2) (\xi + 1) \right] \tag{5.270}
\]

\[
\frac{dh_0^0 (\xi)}{d \xi} = \frac{h}{8} \left[ (\xi - 1)^2 + 2 (\xi + 1) (\xi - 1) \right] \tag{5.271}
\]

\[
\frac{dh_1^1 (\xi)}{d \xi} = \frac{h}{8} \left[ (\xi + 1)^2 + 2 (\xi - 1) (\xi + 1) \right] \tag{5.272}
\]

and the second derivatives are

\[
\frac{d^2 h_0^0 (\xi)}{d^2 \xi} = \frac{1}{4} \left[ 2 (\xi - 1) + 2 \{(\xi + 2) + (\xi - 1)\} \right] \tag{5.273}
\]

\[
\frac{d^2 h_0^0 (\xi)}{d^2 \xi} = -\frac{1}{4} \left[ 2 (\xi + 1) + 2 \{(\xi - 2) + (\xi + 1)\} \right] \tag{5.274}
\]

\[
\frac{d^2 h_0^0 (\xi)}{d^2 \xi} = \frac{h}{8} \left[ 2 (\xi - 1) + 2 \{(\xi + 1) + (\xi - 1)\} \right] \tag{5.275}
\]

\[
\frac{d^2 h_1^1 (\xi)}{d^2 \xi} = \frac{h}{8} \left[ 2 (\xi + 1) + 2 \{(\xi - 1) + (\xi + 1)\} \right]. \tag{5.276}
\]

For terms such as \( L_0 (h_0^0 (\xi)) \) we have, using the definitions presented above,
The matrix equation that one obtains after evaluating Eqs. (5.277) through (5.280) at the two Gauss Legendre points, that is at \( \xi = \pm 0.577 \) (see table on pg. 189) and at the end points \( x = 0 \) (where \( \xi = -1 \) in the first element) and \( x = 2 \) (where \( \xi = 1 \) in the second element) is

\[
\begin{bmatrix}
5.5 & 0.5 & 1.5 & 0 \\
-4.52 & -0.628 & -0.345 & 0 \\
0 & -4.02 & -4.52 & -0.628 \\
0 & 2.01 & 0.761 & 0.0448
\end{bmatrix}
\begin{bmatrix}
\frac{du_0}{dx} \\
u_1 \\
\frac{du_1}{dx} \\
u_2
\end{bmatrix}
= 
\begin{bmatrix}
5.5 \\
-2.52 \\
0 \\
0
\end{bmatrix}.
\]

(5.281)

The solution to this problem is

\[
\begin{bmatrix}
\frac{du_0}{dx} \\
u_1 \\
\frac{du_1}{dx} \\
u_2
\end{bmatrix}
= 
\begin{bmatrix}
-0. & -2.8 \\
-0.804 & -0.976 \\
-1.22 & -1.22
\end{bmatrix}
\]

(5.282)

and it is compared with the analytical solution in the following table:

<table>
<thead>
<tr>
<th>Function</th>
<th>Location</th>
<th>Analytical</th>
<th>Numerical</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{du_0}{dx} )</td>
<td>( x = 0 )</td>
<td>-2.80</td>
<td>-2.80</td>
<td>-0.000</td>
</tr>
<tr>
<td>( u_1 )</td>
<td>( x = 1 )</td>
<td>-0.801</td>
<td>-0.804</td>
<td>-0.003</td>
</tr>
<tr>
<td>( \frac{du_1}{dx} )</td>
<td>( x = 1 )</td>
<td>-0.965</td>
<td>-0.976</td>
<td>-0.011</td>
</tr>
<tr>
<td>( u_2 )</td>
<td>( x = 2 )</td>
<td>-1.22</td>
<td>-1.22</td>
<td>0</td>
</tr>
</tbody>
</table>
5.6 Chapter Summary

In this chapter we introduced the key elements underlying the theory that we will use for all of the numerical methods presented in this book, except finite differences. The key concepts are the weighted residual formalism and the use of polynomial approximation theory to represent the unknown and sought-after solutions to differential equations and their associated boundary conditions. By selecting different definitions of the weighting functions in the weighted residual formulations, the Galerkin finite element method, the finite volume method and the collocation method are defined. The programming strategy that is generally used to implement these methods is based upon the concept of first evaluating the integrals that are required by the formulation at the element level using a local $0 \leq \chi \leq 1$ coordinate system. Thus there is generated an element coefficient matrix for each element. The information from each element is then collected to form the global coefficient matrix that is actually used to solve the problem. To tie the various methods together, the Galerkin finite element and finite volume methods are interpreted in terms of finite difference formulae. Examples for each numerical approach are presented and their results compared to those generated analytically.

5.7 Problems

1. Consider

\[
\frac{du(x)}{dx} + Q = 0 \quad x \in [0, 1] \\
u(0) = 1 \\
Q = 2.
\]

Calculate the value at $u(1)$ using the finite volume method and one element.

2. Consider the equation

\[
\frac{d^2 u(x)}{dx^2} = 0 \quad x \in [0, 1] \\
u(0) = 1 \\
\frac{du(x)}{dx} \bigg|_{x=1} = 0.
\]

Use the finite element method with three nodes to solve for $x = 0.5$.

3. Assume we have an integral of the form

\[
\int_0^3 \frac{d\phi(L)}{dx} dx.
\]

Transform this integral into local $0 \leq \chi \leq 1$ coordinates and evaluate the integral.
4. Consider the equation
\[
\frac{du(x)}{dx} + u(x) - 1 = 0 \quad x \in [0, 1] \quad (5.290)
\]
\[
u(0) = 2.
\]

Solve this equation using both the finite difference method and the finite volume or subdomain method and compare the results for the analytical solution
\[
u(x) = 1 + \exp(-x). \quad (5.291)
\]

You will use 17 nodes. In the case of the finite difference method represent the \(u(x)\) term as \(u_i\), and a backward difference approximation. Alternatively, you could imagine you are standing in the middle between two nodes and define \(u_{i+\frac{1}{2}} = \frac{u_i + u_{i-1}}{2}\) as the approximation. You can also use a second-order finite difference approximation for the first derivative.

To do the finite volume model:

1. (a) create the element coefficient matrix; it will be \(2 \times 2\) and the same for every element.
   (b) create the global matrix by using a strategy that relates the global node numbers to the local node numbers
   (c) impose the boundary conditions
   (d) solve the problem.

Now, plot the solutions for the two finite difference methods, subdomain method and the analytic solution. Compare the relative errors of the three methods and discuss why they are different.
5. Consider the equation

\[ \frac{d^2 u(x)}{dx^2} = 1. \]  \hspace{1cm} (5.292)

Develop the Galerkin finite element representation of this equation at the node \( i \) in Fig. 5.17. Do the following:

(a) write the weighted residual approximation using the linear Lagrange polynomial as the weighting function

(b) substitute the above equation replacing \( u(x) \) by \( \hat{u}(x) \) into the weighted residual formulation

(c) apply integration by parts to the second-order term yielding an integral of the product of two first-order terms and two boundary terms (ignore the boundary terms)

(d) introduce the definitions for \( \hat{u}(x) \), that is introduce the Lagrange polynomial expansion.

(e) write the element matrix equations with the coefficients in the matrix presented in symbolic form, that is show them as integrals with appropriate integration limits.

(f) assemble the global matrix equation using the information generated in the preceding step.

Note: Do not do anything in local coordinates and do not do the integrations

\[ \begin{array}{c}
\text{Figure 5.17: Nodal arrangement used to solve problem 5.7.} \\
\end{array} \]

6. In Section 5.3 on page 102 we developed the solution to the problem

\[ \frac{d^2 u(x)}{dx^2} + \frac{du(x)}{dx} + \frac{u(x)}{2} = 0 \quad x \in [0, 3] \]

\[ u(0) = 1 \]

\[ \left. \frac{du(x)}{dx} \right|_3 = 0 \]

with analytic solution

\[ u(x) = \exp\left(-x/2\right) \left[ \cos\left(x/2\right) + \tan\left(\pi/4 + 3/2\right) \sin\left(x/2\right) \right]. \]

The problem is to determine how your numerical solution to this problem changes as you change the value of \( h \). Please do the following:

(a) Create a Matlab code that solves the above problem using the Galerkin finite element formulation shown in Section 5.3;

(b) Solve the same problem using \( h = h/2, h/3, h/4, h/5 \); you should have a common node at \( x = 1.0 \) and 2.0;
(c) Determine the error at $x = 1$ and $x = 2$ for each spacing;

(d) Plot the log of the error at $x = 1.0$ and $2.0$ versus the log of the spacing and calculate the slope. The slope of this line relates to the order $n$ of the error of the approximation, that is $O(h^n)$;

(e) What is the order of the error, that is, what is $n$.

(f) Explain why a cubic Lagrange polynomial cannot be used for the collocation solution to a second order equation and a cubic Hermite can.

(g) What are the advantages and disadvantages of using two finite elements with three linear basis functions versus one element using three quadratic basis functions.

(h) The finite volume method is said to exactly preserve local mass conservation when solving the time independent convective-diffusion equation, that is

$$ \frac{d^2u}{dx^2} - v \frac{du}{dx} = 0 \quad (5.293) $$

The reason for this lies in the fact that the finite volume boundaries have the same value of the derivative approximation $\frac{du}{dx}$ on each side. Show why this is the case.
Bibliography


Chapter 6

Initial Boundary-Value Problems

6.1 Introduction

In this chapter we are interested in solving problems that have both a spatial and temporal component. As a point of departure, let us consider the equation

\[
\frac{\partial}{\partial t}u(x, t) + \frac{\partial}{\partial x}u(x, t) = 0 \quad t \in [t_0 = 0, \quad t_N = T) \quad x \in [A, \quad B].
\]  

(6.1)

A typical initial condition for this equation would be to specify the state of the system at some time designated as zero such as

\[
u(x, 0) = u_0(x)
\]

and a typical boundary condition that could be a function of time as

\[
u(0, t) = u_1(t) \quad t \geq 0.
\]

(6.3)

6.2 Two Dimensional Polynomial Approximations

To achieve this goal we need to extend our early Lagrange polynomial approximation strategy to consider two dimensions, namely space and time. To this end we define the following quantities:

\[
x \in [x_0 = a, \quad x_n = b]
\]

(6.4)

\[
t \in [t_0 = c, \quad t_m = d]
\]

(6.5)

\[n \equiv \text{degree of the Lagrange polynomial in space, that is } \ell_i(x)^n\]

(6.6)

\[m \equiv \text{degree of the Lagrange polynomial in time, that is } \ell_j(t)^m\]

(6.7)

\[x = i\Delta x\]

(6.8)

\[t = j\Delta t\]

(6.9)

Note that this means that to define the Lagrange polynomial \(\ell_i^n(x)\) there are \(n + 1\) nodes in the \(x\) direction between \(a\) and \(b\). Similarly, to define the Lagrange polynomial in time \(\ell_j^m(t)\) there are \(m + 1\) nodes in the time dimension between \(c\) and \(d\). The
appropriate approximation for $f(x,t)$ is obtained by taking the product of these two functions, that is

$$f(x,t) = \sum_{i=0}^{n} \sum_{j=0}^{m} \ell_i^n(x) \ell_j^m(t) f(x_i,t_j) + E_{nm}$$

(6.10)

where $E_{nm}$ is the error term. Let

$$\hat{f}(x,t) = \sum_{i=0}^{n} \sum_{j=0}^{m} \ell_i^n(x) \ell_j^m(t) f(x_i,t_j).$$

(6.11)

A pictorial representation of $\ell_i^n(x) \ell_j^m(t)$ is found in Fig. 6.1 for the case of $n = i$ and $m = j$.

![Figure 6.1: A conceptual representation of the function $\ell_i^n(x) \ell_j^m(t)$. The dots represent nodes and the maximum function value is unity.](image)

Recall the definition

$$p_n(x) = \Pi_{i=0}^{i=n} (x - x_i)$$

(6.12)

where $\Pi_i$ is the product operator. Using this notation we can write the form of the error $E_{nm}$ in Eq. (6.10) as (see [1])

$$E_{nm}(x,t) = \frac{p_n(x)}{(n+1)!} \frac{\partial^{(n+1)} f(x, t)}{\partial x^{(n+1)}} + \frac{p_m(t)}{(m+1)!} \frac{\partial^{(m+1)} f(x, t)}{\partial t^{(m+1)}}$$

$$- \frac{p_n(x) p_m(t)}{(m+n+2)!} \frac{\partial^{(n+1)} f(x, t)}{\partial x^{(n+1)}} \frac{\partial^{(m+1)} f(x, t)}{\partial t^{(m+1)}} f(\xi', \eta')$$

(a < x, \xi' < b; \ c < t, \eta' < d)

(6.13)

### 6.2.1 Example of a Two Dimensional Polynomial Approximation

Now we consider an example that illustrates the approximation of a function in two space dimensions using Eq. (6.11). We will consider the representation of $\exp(x,t)$
6.3. **FINITE DIFFERENCE APPROXIMATION**

over $[-1,1] \times [-1,1]$ with $m = n = 2$ and $\Delta x = \Delta t = 1$ (see Fig. 6.2). From Eq. (6.11) we have for the representation of $\exp(x,t)$

$$\exp(x,t) = \sum_{i=0}^{2} \sum_{j=0}^{2} \exp(x_i, t_j) \ell^2_i(x) \ell^2_j(t) + E_{2,2} \quad (6.14)$$

where

$$\ell^2_0(x) = \frac{1}{2}x(x-1) \quad (6.15)$$
$$\ell^2_1(x) = -(x+1)(x-1) \quad (6.16)$$
$$\ell^2_2(x) = \frac{1}{2}x(x+1) \quad (6.17)$$

The functions $\ell^2_j(t)$ have an analogous form in the time domain.

If we neglect the error term and substitute the definitions of the Lagrange polynomials into Eq. (6.14) we obtain

$$\exp(x,t) =$$

$$\frac{1}{2}x(x-1) \left[ \frac{1}{2}t(t-1) \exp(-1,-1) - (t+1)(t-1) \exp(-1,0) + \frac{1}{2}t(t+1) \exp(-1,1) \right] -$$

$$2x \left[ \frac{1}{2}t(t-1) \exp(0,-1) - (t+1)(t-1) \exp(0,0) + \frac{1}{2}t(t+1) \exp(0,1) \right] +$$

$$\frac{1}{2}x(x+1) \left[ \frac{1}{2}t(t-1) \exp(1,-1) - (t+1)(t-1) \exp(1,0) + \frac{1}{2}t(t+1) \exp(1,1) \right] \quad (6.18)$$

where the $\exp(x,t)$ values in the expansion are those found at the nodes. Thus Eq. (6.19) allows the determination of the value of $\exp(x,t)$ anywhere in the area $x \in [-1,1], y \in [-1,1]$ based upon the point values at the nodes as indicated in Fig. 6.2.

6.3 **Finite Difference Approximation**

Following the logic we used earlier for one-dimensional problems, we now formulate the approximation of a derivative of $f(x,t)$ defined in two dimensions, for example, $\frac{\partial}{\partial x} f(x,t)$. Consider the segment of the grid shown in gray in Fig. 6.3. This area can be expressed as

$$[a,b] = [x_i, x_{i+1}] \quad (6.20)$$
$$[c,d] = [t_j, t_{j+1}] \quad (6.21)$$

The approximation of $\frac{\partial}{\partial x} f(x_i, t_j)$ over this rectangular area can be obtained directly by differentiation of Eq. (6.10). We obtain

$$\frac{\partial}{\partial x} f(x,t) = \sum_{i=0}^{m} \sum_{j=0}^{n} \frac{\partial \ell^m_i(x)}{\partial x} \ell^m_j(t) f(x_i, t_j) + \frac{\partial}{\partial x} E_{nm} \quad (6.22)$$
which, when customized for our problem, becomes

\[
\frac{\partial f(x, t)}{\partial x} \bigg|_{x_i, t_j} = f(x_i, t_j) \frac{\partial \ell_i(x)}{\partial x} \bigg|_{x_i} \ell_j(t_j) + f(x_{i+1}, t_j) \frac{\partial \ell_{i+1}(x)}{\partial x} \bigg|_{x_i} \ell_j(t_j) \\
+ f(x_{i+1}, t_{j+1}) \frac{\partial \ell_{i+1}(x)}{\partial x} \bigg|_{x_i} \ell_{j+1}(t_j) \\
+ f(x_{i+1}, t_{j+1}) \frac{\partial \ell_{i+1}(x)}{\partial x} \bigg|_{x_i} \ell_{j+1}(t_j) + \frac{d}{dx}(E_{i,j}). \quad (6.23)
\]

Substitution of the definitions of the Lagrange polynomials yields

\[
\frac{\partial f(x, y)}{\partial x} \bigg|_{x_i, t_j} = f(x_i, t_j) \frac{1}{x_i - x_{i+1}} \frac{t_j - t_{j+1}}{t_j - t_{j+1}} + f(x_{i+1}, t_j) \frac{1}{x_{i+1} - x_i} \frac{t_j - t_{j+1}}{t_j - t_{j+1}} \\
+ f(x_{i+1}, t_{j+1}) \frac{1}{x_{i+1} - x_i} \frac{t_j - t_{j+1}}{t_{j+1} - t_j} + \frac{d}{dx}(E_{i,j}). \quad (6.24)
\]
Upon substitution of nodal coordinates we get (note that $p_1 = (x - x_0)(x - x_1)$)

$$\frac{\partial}{\partial x} f(x_i, t_j) =$$
$$\frac{1}{\Delta x} [f(x_{i+1}, t_j) - f(x_i, t_j)] +$$
$$\frac{p_1(x_i) \partial^3}{2! \partial x^3} f(\xi, t_j) + \left[ \frac{(x_i - x_i) + (x_i - x_i + 1)}{\Delta x} \right] \frac{\partial^2}{\partial x^2} f(\xi, t_j) +$$
$$\frac{p_1(t_j) \partial^2}{2! \partial x \partial t} f(x_i, \eta) + \left[ \frac{(x_i - x_i) + (x_i - x_i + 1)}{\Delta t} \right] \frac{\partial^2}{\partial x^2} \frac{\partial}{\partial t} f(x_i, \eta) -$$
$$\frac{1}{4!} \left[ p_1(x_i) p_1(t_j) \right] \frac{\partial^3}{\partial x^3} \frac{\partial^2}{\partial t^2} f(\xi', \eta') \quad (6.25)$$

After defining $\Delta x \equiv x_{i+1} - x_i$ and $\Delta t = t_{j+1} - t_j$ we can write

$$\frac{\partial}{\partial x} f(x_i, t_j) = \frac{1}{\Delta x} [f(x_{i+1}, t_j) - f(x_i, t_j)] + O(\Delta x). \quad (6.26)$$

Similarly one can show that

$$\frac{\partial}{\partial t} f(x_i, t_j) = \frac{1}{\Delta t} [f(x_i, t_{j+1}) - f(x_i, t_j)] + O(\Delta t). \quad (6.27)$$

A computational molecule (or template) of the approximation shown in Eqs. (6.26) and (6.27) is found in Fig. (6.3). The gray area is the area represented by this molecule.

![Figure 6.3: Computational molecule for Eqs. 6.26 and 6.27 shown in grey.](image)
6.3.1 Example of Implicit First-Order Accurate Finite Difference Calculation

Consider the type equation presented above, namely

\[ \frac{\partial u(x, t)}{\partial x} + \frac{\partial u(x, t)}{\partial t} = 0 \]  

(6.28)

and impose the following initial and boundary conditions

\[ u(x, 0) = 0 \]  

(6.29)

\[ u(0, t) = 1. \]  

(6.30)

Now evaluate Eq. (6.24) at location \( i + 1, j + 1 \), that is evaluate \( \frac{\partial}{\partial x} u(x_{i+1}, t_{j+1}) \). Also obtain \( \frac{\partial}{\partial t} u(x_{i+1}, t_{j+1}) \). Let \( i = 0, 1, 2 \) and \( j = 0, 1, 2 \). Also assume \( \Delta x = \Delta t = 1 \). For convenience, define \( \rho = \frac{\Delta t}{\Delta x} = 1 \). The finite difference expression for Eq. (6.28) using the node \( x_{i+1}, t_{j+1} \) as the point of reference is

\[ \frac{u_{i+1,j+1} - u_{i,j+1}}{\Delta t} + \frac{u_{i+1,j+1} - u_{i,j+1}}{\Delta x} = 0. \]  

(6.31)

The computational molecule for this expression is found in Fig. 6.5. Rearranging Eq. (6.31) we obtain

\[ u_{i+1,j+1} = \frac{1}{1+\rho} u_{i+1,j} + \frac{\rho}{1+\rho} u_{i,j+1} \]

\[ = \frac{1}{2} u_{i+1,j} + \frac{1}{2} u_{i,j+1}. \]  

(6.32)

Time Step One

Consider the nodal arrangement in Fig. 6.6. Along the \( x \) axis the values are known for the 0 level time step, \( t = j \Delta t = 0 \) from the initial conditions. Thus, there is a value of 0 associated with each of these nodes. For the left-hand boundary, \( x = 0 \), the values are known for all time by the boundary condition; therefore the value at \( x = 0 \) for all time is 1. So, the first value we actually need to calculate is at node 1, 1. The equation is
Figure 6.5: Computational molecule for implicit backward in time and backward in space first derivative approximations.

\[
\begin{align*}
    u_{1,1} &= \frac{1}{2} u_{1,0} + \frac{1}{2} u_{0,1} \\
    &= \frac{1}{2} (0) + \frac{1}{2} (1) \\
    &= \frac{1}{2}.
\end{align*}
\]  

(6.33)

Now that we know the value \( u_{1,1} = \frac{1}{2} \) we can proceed to node 2,1. The equation at node 2,1 is

\[
\begin{align*}
    u_{2,1} &= \frac{1}{2} u_{2,0} + \frac{1}{2} u_{1,1} \\
    &= \frac{1}{2} (0) + \frac{1}{2} \left( \frac{1}{2} \right) \\
    &= \frac{1}{4}.
\end{align*}
\]

(6.34)

This completes the calculations for time step one.

**Time Step Two**

We are now on the third row of the computational molecule at location \( t = 2 \). The first unknown value moving left to right along the line at \( t = 2 \) is

\[
\begin{align*}
    u_{1,2} &= \frac{1}{2} u_{1,1} + \frac{1}{2} u_{0,2} \\
    &= \frac{1}{2} \left( \frac{1}{2} \right) + \frac{1}{2} (1) \\
    &= \frac{3}{4}.
\end{align*}
\]

(6.35)

and once the value at \( u_{1,2} \) is known we can proceed to \( u_{2,2} \), where the equation is

\[
\begin{align*}
    u_{2,2} &= \frac{1}{2} u_{2,1} + \frac{1}{2} u_{1,2} \\
    &= \frac{1}{2} \left( \frac{1}{4} \right) + \frac{1}{2} \left( \frac{3}{4} \right) \\
    &= \frac{1}{2}.
\end{align*}
\]

(6.36)
We have now completed two time steps. The solution for each time step is shown graphically in Fig. 6.7.

Figure 6.7: Solution to finite difference approximation after two time steps have been executed.

### 6.3.2 Example of Second Order Accurate Finite Difference Approximation in Space

We will now consider the possibility of representing the space derivative in the example problem using a different spatial approximation. Consider once again the type equation presented above, namely

\[
\frac{\partial u(x,t)}{\partial t} + \frac{\partial u(x,t)}{\partial x} = 0 \quad x \in [0,2], \quad t \in [0,2]
\]  

(6.37)
and impose the following initial and boundary conditions

\[ u(x, 0) = 0 \quad (6.38) \]

\[ u(0, t) = 1. \quad (6.39) \]

We now need to select a suitable finite difference approximation to this equation. Let us use the following centered in space and backward in time approximations

\[ \frac{u_{i,j+1} - u_{i,j}}{\Delta t} + \frac{u_{i+1,j+1} - u_{i-1,j+1}}{2\Delta x} = 0 \quad (6.40) \]

or, upon rearranging so the unknown information is on the left hand side of the equation and the known information is on the right,

\[ \frac{u_{i,j+1}}{\Delta t} + \frac{u_{i+1,j+1} - u_{i-1,j+1}}{2\Delta x} = \frac{u_{i,j}}{\Delta t}. \quad (6.41) \]

The template we wish to use for this equation is shown in Fig. 6.8, where the unknown values are located at filled in circles and the known information is at the open circles.

![Figure 6.8: Computational molecule for an implicit first-order accurate finite difference calculation in time and a second-order accurate approximation in space.](image)

**Step One**

Write Eq. (6.41) for the special case of \( \Delta t = 1 \) and \( \Delta x = 1 \) as shown in Fig. 6.6 which is

\[ \frac{u_{1,1}}{1} + \frac{u_{2,1} - u_{0,1}}{2(1)} = \frac{u_{1,0}}{1}. \quad (6.42) \]

Now we use the fact that the initial condition is given as \( u(x, 0) = 0 \) which implies that \( u_{1,0} = 0 \). Because the boundary condition is \( u(0, t) = 1 \) we have \( u_{0,1} = 1 \). Introducing these conditions into Eq. (6.42) we obtain

\[ \frac{u_{1,1}}{1} + \frac{u_{2,1} - 1}{2(1)} = \frac{0}{1}. \quad (6.43) \]

We cannot write an equation for node 2,1 using Eq. (6.41) because there is no node to the right of node 2,1 which would be the \( i+1 \) node in Eq. (6.41). Thus Eq.
(6.43) has two unknowns. How can we solve this equation? One strategy is to use a backward in space approximation for node 2,1. A suitable expression is

\[
\frac{u_{2,1} - u_{2,0}}{\Delta t} + \frac{u_{2,1} - u_{1,1}}{\Delta x} = 0
\]

which for our specific case becomes

\[
\frac{u_{2,1} - u_{2,0}}{1} + \frac{u_{2,1} - u_{1,1}}{1} = 0.
\]

This would provide two equations in two unknowns which can be solved. This is an implicit approximation. An implicit approximation requires that one solve an equation involving both the current state of the system and the later one. This requires the solution of a system of equations, and in this case the set of equations can be written in matrix form as

\[
\begin{bmatrix}
-\frac{1}{\Delta x} & \frac{1}{\Delta x} & 0 \\
-\frac{1}{2\Delta x} & 0 & \frac{1}{2\Delta x} \\
0 & -\frac{1}{\Delta x} & \frac{1}{\Delta x}
\end{bmatrix}
\begin{bmatrix}
u(0,0) \\
u_{1,1} \\
u_{2,1}
\end{bmatrix}
\]

\[
= -\begin{bmatrix}
\frac{1}{\Delta t} & 0 & 0 \\
1 & \frac{1}{\Delta t} & 0 \\
1 & 0 & \frac{1}{\Delta t}
\end{bmatrix}
\begin{Bmatrix}
u(0,0) \\
u_{1,1} \\
u_{2,1}
\end{Bmatrix}
- \begin{Bmatrix}
u(1,0) \\
u_{1,0} \\
u_{2,0}
\end{Bmatrix}
\]

where the term identified as \( A \) represents the space derivative and the term identified as \( B \) the time derivative. Note that some unknown variables have been replaced by known values represented by initial and boundary conditions.

Since \( u(0,t) \) is known to be 1, the first column can be multiplied by this value and the first row and column now contain no useful information and can be eliminated to give

\[
\begin{bmatrix}
0 & \frac{1}{\Delta x} \\
-\frac{1}{\Delta x} & \frac{2\Delta x}{\Delta x}
\end{bmatrix}
\begin{bmatrix}
u_{1,1} \\
u_{2,1}
\end{bmatrix}
\]

\[
= -\begin{bmatrix}
\frac{1}{\Delta t} & 0 & 0 \\
1 & \frac{1}{\Delta t} & 0 \\
1 & 0 & \frac{1}{\Delta t}
\end{bmatrix}
\begin{Bmatrix}
u_{1,1} \\
u_{2,1}
\end{Bmatrix}
- \begin{Bmatrix}
u_{1,0} \\
u_{2,0}
\end{Bmatrix}
+ \begin{Bmatrix}
\frac{1}{2\Delta x} \\
0
\end{Bmatrix}
\]

\[
\text{(6.48)}
\]

The term \( C \) now contains the information about the boundary condition.

Now collect all the information regarding the unknown vector \( \begin{bmatrix} u_{1,1} \\ u_{2,1} \end{bmatrix} \) to the left-hand side of the equation to yield

\[
\begin{Bmatrix}
0 & \frac{1}{\Delta x} \\
-\frac{1}{\Delta x} & \frac{2\Delta x}{\Delta x}
\end{bmatrix}
\begin{bmatrix}
u_{1,1} \\
u_{2,1}
\end{bmatrix}
\]

\[
= -\begin{bmatrix}
\frac{1}{\Delta t} & 0 & 0 \\
1 & \frac{1}{\Delta t} & 0 \\
1 & 0 & \frac{1}{\Delta t}
\end{bmatrix}
\begin{Bmatrix}
u_{1,1} \\
u_{2,1}
\end{Bmatrix}
+ \begin{Bmatrix}
\frac{1}{2\Delta x} \\
0
\end{Bmatrix}
\]

\[
\text{(6.47)}
\]

We now substitute the specific values for \( \Delta x = 1 \) and \( \Delta t = 1 \) to give
6.4. Stability of Finite Difference Approximations

Several approaches exist to determine the stability of a numerical scheme that involves derivatives in space and time. The approach we will employ is called the Matrix Method. Consider the equation

\[
\frac{\partial u(x, t)}{\partial t} - \frac{\partial^2 u(x, t)}{\partial x^2} = 0 \quad x \in [0, L] \quad t \in [0, \infty)
\]  

(6.56)

with boundary and initial conditions (note that this is second order in space, so we will need two boundary conditions in \( x \) as well as an initial condition).
In this development we will employ, as an example, an explicit forward difference approximation in time and a centered finite difference approximation in space. Introduction of these approximations yields the finite difference equation

\[
\frac{1}{\Delta t} [u(x, t_{j+1}) - u(x, t_j)] = \frac{1}{\Delta x^2} [u(x_{i+1}, t_j) - 2u(x_i, t_j) + u(x_{i-1}, t_j)].
\]  

(6.60)

To simplify notation let

\[
r = \frac{\Delta t}{\Delta x^2}
\]

(6.61)

and rewrite the finite difference formula as

\[
u_{i,j+1} = ru_{i+1,j} + (1 - 2r) u_{i,j} + ru_{i-1,j} \quad i = 0, 1, ..., N.
\]

(6.62)

Now we write the matrix form of Eq. (6.62) for the \(j+1\) time step. After imposing the boundary conditions (remember the boundary conditions are zero so no term appears on the right-hand side to identify them) we have

\[
\begin{bmatrix}
u_{1,j+1} \\
u_{2,j+1} \\
u_{3,j+1} \\
\vdots \\
u_{N-1,j+1}
\end{bmatrix} =
\begin{bmatrix}
(1 - 2r) & r & 0 & 0 & 0 & 0 \\
r & (1 - 2r) & r & 0 & 0 & 0 \\
0 & r & (1 - 2r) & r & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
u_{1,j} \\
u_{2,j} \\
u_{3,j} \\
\vdots \\
u_{N-1,j}
\end{bmatrix}
\]

(6.63)

We can write Eq. (6.63) symbolically as

\[\{u\}_{j+1} = [A] \{u\}_j.\]

(6.64)

Assume an exact solution, free of round-off error is given by \(\{\tilde{u}\}_{j+1}\). Note that this vector contains the values of \(u_{i,j+1}\) at all nodes, except the Dirichlet boundary nodes, at time \((j + 1) \Delta t\). Since \(\{\tilde{u}\}_{j+1}\) is a solution to our equation we can write

\[\{\tilde{u}\}_{j+1} = [A] \{\tilde{u}\}_j.\]

(6.65)

Now subtract Eq. (6.65) from Eq. (6.64) to obtain a measure of the error or difference between the exact and computed values, that is

\[\{u - \tilde{u}\}_{j+1} = [A] \{u - \tilde{u}\}_j.\]

(6.66)

Next define this error as

\[\{\varepsilon\}_{j+1} = \{u - \tilde{u}\}_{j+1}.\]

(6.67)

We can now rewrite Eq. (6.66) as

\[\{\varepsilon\}_{j+1} = [A] \{\varepsilon\}_j.\]

(6.68)
6.4. STABILITY OF FINITE DIFFERENCE APPROXIMATIONS

Assuming no further errors are committed, other than the round-off errors that exist at \( u_0 \), we have

\[
\{\varepsilon\}_1 = [A] \{\varepsilon\}_0 \\
\{\varepsilon\}_2 = [A] \{\varepsilon\}_1 = [A]^2 \{\varepsilon\}_0 \\
\vdots \\
\{\varepsilon\}_j = [A]^j \{\varepsilon\}_0
\]

which illustrates how the error \( \{\varepsilon\}_j \) changes as the solution advances through time.

In the next step we analyze the behavior of this scheme through the use of eigenvalues. Assume: (1) there exist \( N - 1 \) different eigenvalues of \( [A] \), that is, \( \lambda_s, s = 1, 2, \ldots, N - 1 \), and (2) there exist a set of linearly independent eigenvectors, \( \{v\}_s \), \( s = 1, \ldots, N - 1 \). Using these eigenvectors as a kind of basis, we can write

\[
\{\varepsilon\}_0 = \sum_{s=1}^{N-1} c_s \{v\}_s
\]

where \( c_s, s = 1, \ldots, N - 1 \) are numbers. Note also that by definition

\[
[A] \{v\}_s = \lambda_s \{v\}_s.
\]

Now substitute Eq. (6.72) in Eq. (6.71) to obtain

\[
\{\varepsilon\}_j = [A]^j \{\varepsilon\}_0 = [A]^j \sum_{s=1}^{N-1} c_s \{v\}_s.
\]

We now proceed through a series of apparently unrelated steps in pursuit of a specific result. We begin with the modification of Eq. (6.74), that is

\[
\{\varepsilon\}_j = [A]^j \sum_{s=1}^{N-1} c_s \{v\}_s \\
= [A]^{j-1} [A] \sum_{s=1}^{N-1} c_s \{v\}_s \\
= [A]^{j-1} \sum_{s=1}^{N-1} c_s [A] \{v\}_s.
\]
Note that we can use Eq. (6.73) to modify the last expression, that is

\[
[A]^{j-1} \sum_{s=1}^{N-1} c_s [A] \{v\}_s = [A]^{j-1} \sum_{s=1}^{N-1} c_s \lambda_s \{v\}_s
\]

\[
= [A]^{j-2} [A] \sum_{s=1}^{N-1} c_s \lambda_s \{v\}_s
\]

\[
= [A]^{j-2} \sum_{s=1}^{N-1} c_s \lambda_s [A] \{v\}_s
\]

\[
= [A]^{j-2} \sum_{s=1}^{N-1} c_s \lambda_s^2 \{v\}_s
\]

\[
= [A]^{j-2} \sum_{s=1}^{N-1} c_s \lambda_s \lambda_s \{v\}_s
\]

\[
= [A]^{j-2} \sum_{s=1}^{N-1} c_s \lambda_s \lambda_s \{v\}_s
\]

\[
= [A]^{0} \sum_{s=1}^{N-1} c_s \lambda_s \lambda_s \{v\}_s.
\]

(6.76)

So we have

\[
\{\varepsilon\}_j = [A]^{0} \sum_{s=1}^{N-1} c_s \lambda_s \lambda_s \{v\}_s.
\]

(6.77)

Examination of Eq. (6.77) indicates that the error \(\{\varepsilon\}_j\) will grow if the eigenvalue \(\lambda_s\) is greater than 1.

The matrix \([A]\) is known to have the eigenvalues

\[
\lambda_s = 1 - 4 r \sin^2 \left( \frac{s \pi}{2 N^\prime} \right)
\]

(6.78)

where \(N^\prime\) is the maximum number of nodes. Thus, to assure stability

\[
\lambda_s \leq 1
\]

(6.79)

or

\[
\left| 1 - 4 r \sin^2 \left( \frac{s \pi}{2 N^\prime} \right) \right| \leq 1
\]

(6.80)

which translates into

\[
-1 \leq 1 - 4 r \sin^2 \left( \frac{s \pi}{2 N^\prime} \right) \leq 1.
\]

(6.81)

Because \(\sin^2 \left( \frac{s \pi}{2 N^\prime} \right)\) is always positive, the upper bound is always satisfied. For the lower bound we have

\[
-1 \leq 1 - 4 r \sin^2 \left( \frac{s \pi}{2 N^\prime} \right)
\]

(6.82)

\[
\Rightarrow -2 \leq -4 r \sin^2 \left( \frac{s \pi}{2 N^\prime} \right)
\]

(6.83)

\[
\Rightarrow 2 \geq 4 r \sin^2 \left( \frac{s \pi}{2 N^\prime} \right)
\]

(6.84)

\[
\Rightarrow \frac{2}{4 \sin^2 \left( \frac{s \pi}{2 N^\prime} \right)} \geq r
\]

(6.85)

\[
\Rightarrow r \leq \frac{1}{2 \sin^2 \left( \frac{s \pi}{2 N^\prime} \right)}.
\]

(6.86)
The worst-case scenario is when
\[
\sin^2 \left( \frac{8\pi}{2N} \right) = 1. \tag{6.87}
\]
Then the constraint on stability is, according to Eq. (6.86)
\[
r \leq \frac{1}{2}. \tag{6.88}
\]
Recalling the definition of \( r \) (see Eq. 6.61) we obtain
\[
\frac{\Delta t}{\Delta x^2} \leq \frac{1}{2}. \tag{6.89}
\]
Equation (6.89) states that to assure stability, the time step must be related to the space increment such that \( \Delta t \leq \frac{\Delta x^2}{2} \).

### 6.4.1 Example of Stability

**Case of \( r = 1 \)**

Consider the approximation for
\[
\frac{\partial u(x, t)}{\partial t} - \frac{\partial^2 u(x, t)}{\partial x^2} = 0 \quad x = [0, 1] \quad t = [0, 2/9] \tag{6.90}
\]
provided in Eq. (6.62) for the case of four nodes in space, namely
\[
u_{i,j+1} = ru_{i+1,j} + (1 - 2r)u_{i,j} + ru_{i-1,j} \quad i = 0, 1, 2, 3 \quad x \in [0, 1] \tag{6.91}
\]
with boundary and initial conditions
\[
u(x, 0) = 0 \tag{6.92}
\]
\[
u(0, t) = 1 \tag{6.93}
\]
\[
u(1, t) = 0. \tag{6.94}
\]

The problem is defined in Fig. 6.9.

The template or computational molecule for the finite difference approximation to Eq. (6.90) given by Eq. (6.91) is shown in Fig. 6.10. As the calculations proceed, only the shaded node is unknown, that is the clear nodes are known from the preceding time level.

**Step One (\( j = 0 \))** Substitution of \( j = 0 \) into Eq. (6.91) yields
\[
u_{1,1} = (1) \times u_{2,0} + (1) \times u_{1,0} + (1) \times u_{0,0} = 1 \tag{6.95}
\]
\[
u_{2,1} = (1) \times u_{3,0} + (1) \times u_{2,0} + (1) \times u_{1,0} = 0 \tag{6.96}
\]
where the values under the variable names indicate the current value of those variables.
Gamma Chapter 6. Initial Boundary-Value Problems

Figure 6.9: Nodal array for the finite difference solution to Eqs. (6.90)- (6.94).

Figure 6.10: Computational molecule for the Euler (forward difference) approximation to Eq. 6.90 given by Eq. 6.91.

Step Two \((j = 1)\)

\[
\begin{align*}
u_{1,2} &= (1) \times u_{2,1} - (1) \times u_{1,1} + (1) \times u_{0,1} = 0 \\
u_{2,2} &= (1) \times u_{3,1} - (1) \times u_{2,1} + (1) \times u_{1,1} = 1.
\end{align*}
\]  

\(6.97\)

\(6.98\)

Step Three \((j = 2)\)

\[
\begin{align*}
u_{1,3} &= (1) \times u_{2,2} - (1) \times u_{1,2} + (1) \times u_{0,2} = 2 \\
u_{2,3} &= (1) \times u_{3,2} - (1) \times u_{2,2} + (1) \times u_{1,2} = -1.
\end{align*}
\]  

\(6.99\)

\(6.100\)

The solution for the first three time steps \((j = 0, 1, 2)\) is shown in Fig. 6.11. It is clearly unstable.
6.4. STABILITY OF FINITE DIFFERENCE APPROXIMATIONS

Case of $r = 1/4$

Now we reduce the value of $r$ to $r = 1/4$. The calculations follow:

Step One ($j = 0$)

$$u_{1,1} = \left(\frac{1}{4}\right) \times u_{2,0} + \left(\frac{1}{2}\right) \times u_{1,0} + \left(\frac{1}{4}\right) \times u_{0,0} = \frac{1}{4}$$  \hspace{1cm} (6.101)

$$u_{2,1} = \left(\frac{1}{4}\right) \times u_{3,0} + \left(\frac{1}{2}\right) \times u_{2,0} + \left(\frac{1}{4}\right) \times u_{1,0} = 0.$$  \hspace{1cm} (6.102)

Step Two ($j = 1$)

$$u_{1,2} = \left(\frac{1}{4}\right) \times u_{2,1} + \left(\frac{1}{2}\right) \times u_{1,1} + \left(\frac{1}{4}\right) \times u_{0,1} = \frac{3}{8}$$  \hspace{1cm} (6.103)

$$u_{2,2} = \left(\frac{1}{4}\right) \times u_{3,1} + \left(\frac{1}{2}\right) \times u_{2,1} + \left(\frac{1}{4}\right) \times u_{1,1} = \frac{1}{16}.$$  \hspace{1cm} (6.104)

Step Three ($j = 2$)

$$u_{1,3} = \left(\frac{1}{4}\right) u_{2,2} + \left(\frac{1}{2}\right) u_{1,2} + \left(\frac{1}{4}\right) u_{0,2} = \frac{31}{64}$$  \hspace{1cm} (6.105)

$$u_{2,3} = \left(\frac{1}{4}\right) u_{3,2} + \left(\frac{1}{2}\right) u_{2,2} + \left(\frac{1}{4}\right) u_{1,2} = \frac{1}{8}.$$  \hspace{1cm} (6.106)

The solution for the first three steps is provided in Fig. 6.12.
It is clear from Fig. 6.12 that the solution is stable.

While these two examples do not prove that the constraints on \( r \) are necessary, they do suggest that to be the case.

### 6.4.2 Example Simulation

Let us assume we have the following equation and boundary conditions

\[
D \frac{\partial^2 c(x, t)}{\partial x^2} = \frac{\partial c(x, t)}{\partial t} \tag{6.107}
\]

with

\[
c(0, t) = c_{0,j} = 0 \text{ for all } j
\]

\[
\frac{\partial c(L, t)}{\partial x} = \frac{\partial c_{3,j}}{\partial x} = 0 \text{ for all } j \tag{6.108}
\]

\[
c(x, 0) = c_{i,0} = 0 \text{ for all } i, \ i \neq 0.
\]

A centered in space, backwards in time finite difference expression for this equation is

\[
D_i \frac{c_{i+1,j+1} - 2c_{i,j+1} + c_{i-1,j+1}}{(\Delta x)^2} = \frac{c_{i,j+1} - c_{i,j}}{\Delta t} \tag{6.109}
\]

We assume that there are four nodes equally separated by distance \( \Delta x \). The second through fourth rows of the matrix equation for this problem are obtained using Eq. (6.109) for nodes 1, and 2. The first row is used to accommodate the first type (Dirichlet) boundary condition and the last for the second type (Neumann) boundary condition which is obtained by using a backward difference in space approximation for the expression appearing in Eq. (6.108), that is
Collecting the coefficients multiplying the unknown vector at time \((j + 1) \Delta t\) on the left-hand side of the equation we obtain

\[
\begin{bmatrix}
D_i \left( \frac{1}{\Delta x^2} \right) & D_i \left( \frac{-2}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 \\
0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 \\
0 & 0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 \\
0 & 0 & 0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right)
\end{bmatrix}
\begin{bmatrix}
c_{0,j+1} \\
c_{1,j+1} \\
c_{2,j+1} \\
c_{3,j+1}
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix} - \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\Delta t} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\Delta t} & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
c_{1,j} \\
c_{2,j} \\
0 \\
0
\end{bmatrix}.
\tag{6.110}
\]

where the first and last rows provide for the boundary conditions. Now we begin calculating. Since \(c_{i,j}\) is zero for \(j = 0\) we have

\[
\begin{bmatrix}
D_i \left( \frac{1}{\Delta x^2} \right) & D_i \left( \frac{-2}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 & 0 \\
0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 & 0 \\
0 & 0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 \\
0 & 0 & 0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0
\end{bmatrix}
\begin{bmatrix}
c_{0,1} \\
c_{1,1} \\
c_{2,1} \\
c_{3,1}
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix} - \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\Delta t} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\Delta t} & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
c_{1,0} \\
c_{2,0} \\
0 \\
0
\end{bmatrix}.
\tag{6.112}
\]

The initial conditions can now be employed to replace \(c_{1,0}\) and \(c_{2,0}\) with known values to yield

\[
\begin{bmatrix}
D_i \left( \frac{1}{\Delta x^2} \right) & D_i \left( \frac{-2}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 & 0 \\
0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 & 0 \\
0 & 0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0 & 0 \\
0 & 0 & 0 & D_i \left( \frac{-1}{\Delta x^2} \right) & D_i \left( \frac{1}{\Delta x^2} \right) & 0
\end{bmatrix}
\begin{bmatrix}
c_{0,1} \\
c_{1,1} \\
c_{2,1} \\
c_{3,1}
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix} - \begin{bmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{\Delta t} & 0 & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\Delta t} & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
c_{1,0} \\
c_{2,0} \\
0 \\
0
\end{bmatrix}.
\tag{6.113}
\]

Assume that we solve these equations for time step 1 and proceed to time step 2. The resulting equation is the same as the above but for the updated ones in time concentrations.
Repeating this process one more time we obtain for time step 3

\[
\begin{bmatrix}
\frac{1}{(\Delta x)^2} & -\frac{2}{\Delta t} & \frac{1}{(\Delta x)^2} & 0 \\
0 & \frac{1}{(\Delta x)^2} & 0 & 0 \\
0 & 0 & \frac{1}{\Delta t} & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
c_0,2 \\
c_1,2 \\
c_2,2 \\
c_3,2
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}
- \begin{bmatrix}
0 & 0 & 0 & 0 \\
0 & \frac{1}{\Delta t} & 0 & 0 \\
0 & 0 & 0 & \frac{1}{\Delta t} \\
0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
0 \\
c_{1,1} \\
c_{2,1} \\
0
\end{bmatrix}
\]  
(6.114)

and so on.

### 6.5 Galerkin Finite Element Approximations in Time

In this section we extend the ideas introduced in Section 6.3 to consider a time-dependent Galerkin finite element formulation. We will consider the same equation as 6.107, but assume \( D = 1 \), that is

\[
\frac{\partial u (x, t)}{\partial t} - \frac{\partial^2 u (x, t)}{\partial x^2} = 0 \quad x \in [0, 1]
\]  
(6.116)

with initial and boundary conditions

\[
u (x, 0) = u_I (x) = 0
\]  
(6.117)

\[
u (0, t) = \bar{u}_0 (t)
\]  
(6.118)

\[
u (1, t) = \bar{u}_1 (t).
\]  
(6.119)

We begin with the definition of the approximating function for one element; this is similar to Eq. (5.58) on page 95, but in this case \( \hat{u} (x, t) \) is a function of both space \( x \) and time \( t \). We first represent \( u (x, t) \) as

\[
u (x, t) = \hat{u} (x, t) + E (x, t)
\]  
(6.120)

and then define \( \hat{u} (x, t) \) as follows:

\[
\hat{u} (x, t) = \sum_{j=0}^{n} u_j (t) \ell_j^n (x).
\]  
(6.121)
Note the structure of this expression. Where we have previously had constant coefficients \( u_j \) we now have functions of time \( u_j(t) \). We could have used another strategy in which we represent the time dimension using a Lagrange polynomial \( \ell^n_j(t) \) and our approximating expression would be

\[
\hat{u}(x, t) = \sum_{j=0}^{n} u_j \ell^n_j(t) \ell^n_j(x),
\]

but the strategy presented in Eq. (6.121) is a more commonly used approach.

As earlier, we now apply Galerkin's method of weighted residuals for one element, again using the basis function as the weighting function. We obtain

\[
\int_x \left( \frac{\partial \hat{u}(x, t)}{\partial t} - \frac{\partial^2 \hat{u}(x, t)}{\partial x^2} \right) \ell^n_i(x) \, dx = 0 \quad i = 0, 1, \ldots, n.
\]

Using integration by parts on the second-order derivative results in

\[
\int_x \frac{\partial \hat{u}(x, t)}{\partial x} \frac{\partial \ell^n_i(x)}{\partial x} \, dx + \int_x \frac{\partial \hat{u}(x, t)}{\partial t} \ell^n_i(x) \, dx - \frac{\partial \hat{u}}{\partial x} \ell_i(x) \bigg|_0^1 = 0, \quad i = 0, 1, \ldots, n.
\]

We now substitute Eq. (6.121) into Eq. (6.124) to obtain, once again for one element,

\[
\sum_{j=0}^{n} u_j(t) \int_x \frac{\partial \ell^n_j}{\partial x} \frac{\partial \ell^n_i}{\partial x} \, dx + \sum_{j=0}^{n} \frac{du_j(t)}{dt} \int_x \ell^n_j \ell^n_i \, dx - \frac{\partial \hat{u}}{\partial x} \ell^n_i(x) \bigg|_0^1 = 0, \quad i = 0, 1, \ldots, n.
\]

Examination of Eq. (6.125) reveals that it is similar to earlier Galerkin finite element expressions but for one important detail. Both the function \( u_j(t) \) and the time derivative of this function, \( \frac{du_j(t)}{dt} \), appear in this equation. In other words, we have a set of \( n + 1 \) ordinary differential equations in time. If we assume a value for \( n \), say, \( n = 1 \) for linear Lagrange functions, we can look at multiple elements by changing the notation to

\[
\sum_{j=0}^{N} u_j(t) \int_x \frac{\partial \ell_j}{\partial x} \frac{\partial \ell_i}{\partial x} \, dx + \sum_{j=0}^{N} \frac{du_j(t)}{dt} \int_x \ell_j \ell_i \, dx - \frac{\partial \hat{u}}{\partial x} \ell_i(x) \bigg|_0^1 = 0, \quad i = 0, 1, \ldots, N
\]

where \( N \) is the total number of nodes and \( \ell_j(x) \) is the node-wise defined Lagrange polynomial; that is each \( \ell_j(x) \) ranges over multiple elements (in this case, two). Since Eq. (6.126) now represents a system of elements with \( N + 1 \) equations in \( N + 1 \) unknowns, we can write it in matrix form

\[
[A] \{u(t)\} + [B] \left\{ \frac{du(t)}{dt} \right\} + \{f\} = 0
\]

where the coefficients of matrices \([A]\) and \([B]\) are defined as
where, as we saw earlier,

\[ a_{ij} = \int_{0}^{1} \frac{d\ell_j}{dx} \frac{d\ell_i}{dx} dx \]  

(6.129)

\[ b_{ij} = \int_{0}^{1} \ell_j \ell_i dx \]  

(6.130)

\[ u_i = u_i(t) \]  

(6.131)

Where there is one or two second type or Neumann boundary conditions we also have

\[ f_i = -\frac{d\hat{u}}{dx} \bigg|_{i}^{1} \bigg|_{0} \]  

(6.132)

defined as non-zero in either the first or last row of the vector of known values, \( f_i \), depending upon whether the Neumann condition is specified at the first or last node, respectively. Otherwise \( f_i \) contains information on Dirichlet boundary conditions (such as in the above example).

So the question arises as to how to efficiently treat the time derivative. We will consider two possibilities.

### 6.5.1 Strategy One: Forward Difference Approximation

Let us approximate \( \frac{du}{dt} \) by a forward difference approximation

\[
\left\{ \frac{du(t)}{dt} \right\}_k = \frac{\{u\}_{k+1} - \{u\}_k}{\Delta t}.
\]  

(6.133)
Then substitute Eq. (6.133) into Eq. (6.127) to obtain from

\[
[A] \{u(t)\} + [B] \left\{ \frac{du(t)}{dt} \right\} + \{f\} = 0
\]  
(6.134)

the expression

\[
[A] \{u\}_k + [B] \left\{ \frac{u\}_{k+1} - \frac{u\}_k}{\Delta t} \right\} + \{f\} = 0.
\]  
(6.135)

Note that the spatial approximation involving \([A] \{u\}_k\) is evaluated at the time \(t = k\Delta t\). We now rearrange this equation to give

\[
[B] \{u\}_{k+1} = (\Delta t [A]) \{u\}_k - \Delta t \{f\}.
\]  
(6.136)

Because the coefficients

\[
b_{ij} = \int_{x=0}^{x=1} \ell_j \ell_i dx
\]  
(6.137)

create a tridiagonal \([B]\) matrix, Eq. (6.136) constitutes a system of equations. The solution of the problem is written formally as

\[
\{u\}_{k+1} = [B]^{-1} (\Delta t [A]) \{u\}_k - \Delta t \{f\}.
\]  
(6.138)

An important observation is that although we are using a forward difference equation, which in the case of finite difference approximations yielded one equation in one unknown, here, as noted, we have a matrix equation. This is due to the fact that the matrix \([B]\) is no longer diagonal. In other words, the three elements in each row of the matrix \([B]\) tie adjacent nodes together. The consequence of this from a practical perspective is that, although the method is explicit, with its stability limitations, it is nevertheless necessary to take an inverse of matrix \([B]\) to get a solution.

### 6.5.2 Strategy Two: Backward Difference Approximation

In this approach we employ a backward difference in time approximation

\[
\left. \left\{ \frac{du(t)}{dt} \right\} \right|_{(k+1)} = \frac{\{u\}_{k+1} - \{u\}_k}{\Delta t}.
\]  
(6.139)

Then from Eq. (6.127) we have

\[
[A] \{u(t)\} + [B] \left\{ \frac{du(t)}{dt} \right\} + \{f\} = 0
\]  
(6.140)

which becomes

\[
[A] \{u\}_{k+1} + [B] \left\{ \frac{u\}_{k+1} - \frac{u\}_k}{\Delta t} \right\} + \{f\} = 0.
\]  
(6.141)

We can rearrange this equation to give

\[
(\Delta t [A] + [B]) \{u\}_{k+1} = [B] \{u\}_k - \Delta t \{f\}.
\]  
(6.142)

We can rewrite this expression as

\[
[C] \{u\}_{k+1} = [B] \{u\}_k - \Delta t \{f\}
\]  
(6.143)

where

\[
[C] = (\Delta t [A] + [B])
\]  
(6.144)
One can formally obtain as a solution for \( \{u\}_{k+1} \), that is
\[
\{u\}_{k+1} = [C]^{-1} ([B] \{u\}_k - \Delta t \{f\}).
\] (6.145)

As noted earlier, the formulation that generates Eq. (6.138) is explicit. That is the space derivative is at the old time level \( j \) and the solution is computed at the new time level \( j + 1 \). This formulation is conditionally stable, that is there is a restriction on the value of \( r \equiv \frac{\Delta t}{\Delta x^2} \), although we do not prove it here. On the other hand, the formulation that generates Eq. (6.145) is implicit. The space derivative approximation is at the new time level. It is unconditionally stable.

If the \([B]\) matrix were to be diagonal, that is there are non-zero values only along the diagonal, then, as before, the explicit formulation would be very efficient, involving one equation in one unknown for each row of the matrix equation. However, it is not, so a complete matrix solution is required.

On the other hand, the implicit formulation shown in Eq. (6.145), which requires a complete matrix equation solution, is unconditionally stable and therefore one can move forward in time with larger time steps. In essence, the advantage of the explicit approximation in the finite difference formulation (where we have one equation in one unknown for each node), is normally not transferable to the finite element case. There is a procedure called mass lumping which we discuss later in Chapter 10 on page 273 that allows for a more efficient solution to the explicit formulation equations.

### 6.6 Chapter Summary

In this chapter we introduce the extension of spatial approximations to include time dependence. Using polynomial approximation theory we create finite difference formulæ that address transient one space-dimensional partial-differential equations. This is followed by a discussion of stability including illustrative examples of how an instability can be exhibited. The Galerkin finite element method is then reformulated to consider time-dependent problems in which both forward and backward in time finite difference methods are presented.

### 6.7 Problems

1. Consider the segment of the grid shown in gray in Figure 6.3. This area can be expressed as
\[
[a, b] = [x_i, x_{i+1}]
\]
\[
[c, d] = [t_j, t_{j+1}].
\] (6.146) (6.147)

Obtain a finite difference approximation to \( \frac{\partial}{\partial x} f(x, t_j) \) over this rectangular area. Begin with Eq. (6.10)
\[
\frac{\partial}{\partial t} f(x, t) = \sum_{i=0}^{m} \sum_{j=0}^{m} \frac{\partial \ell_i^n (x)}{\partial t} \ell_j^m (t) f(x_i, t_j) + \frac{\partial}{\partial t} E_{nm}
\] (6.148)

and customize it for this use. Then use the definitions of the Lagrange polynomials to obtain the finite difference form.
2. Consider the type equation
\[ \frac{\partial^2 u(x, t)}{\partial x^2} - \frac{\partial u(x, t)}{\partial t} = 0 \]
and impose the following initial and boundary conditions
\[ u(x, 0) = 0 \quad (6.150) \]
\[ u(0, t) = 1 \quad (6.151) \]
\[ u(2, t) = 0. \quad (6.152) \]

Now approximate Eq. (6.28) at location \(i, j+1\) using a backward difference in time. Let \(i = 0, 1, 2,\) and \(j = 0, 1, 2,\) where \(x = i\Delta x\) and \(t = j\Delta t.\) Also assume \(\Delta x = \Delta t = 1\) such that \(\rho = \frac{\Delta t}{\Delta x^2} = 1.\) The finite difference expression for Eq. (6.28) using the node \(x_i, t_{j+1}\) as the point of reference is
\[ \frac{u_{i,j+1} - u_{i,j}}{\Delta t} = \frac{u_{i+1,j+1} - 2u_{i,j+1} + u_{i-1,j+1}}{\Delta x^2} = 0. \quad (6.153) \]

The computational molecule for this expression is found in Figure 6.13. Calculate the values of \(u\) at time levels \(\Delta t\) and \(2\Delta t.\) Here you have an ambiguity at \(u_{i-1,j}.\) Assume the initial condition as the boundary condition at this node to resolve this. Should this scheme be stable? How do you know?

3. Consider the type equation
\[ \frac{\partial^2 u(x, t)}{\partial x^2} - \frac{\partial u(x, t)}{\partial t} = 0 \]

(a) obtain a finite difference form of Eq. (6.154) at location \(i, j + 1/2,\) that is, evaluate \(\frac{\partial^2}{\partial x^2} u(x_i, t_{j+1/2})\)

(b) obtain the finite difference form of \(\frac{\partial}{\partial t} u(x_i, t_{j+1/2}).\)

Write the finite difference equation at the point \((i, j + 1/2)\) (see Fig. 6.14).

4. Consider the equation
\[ \frac{\partial u(x, t)}{\partial t} - \frac{\partial^2}{\partial x^2} u(x, t) = 0 \]
with boundary and initial conditions
164

CHAPTER 6. INITIAL BOUNDARY-VALUE PROBLEMS

If you employ an implicit backward difference approximation in time and a centered finite difference approximation in space, you will obtain

\[
\frac{1}{\Delta t} [u(x_i, t_{j+1}) - u(x_i, t_j)] =
\frac{1}{\Delta x^2} [u(x_{i+1}, t_{j+1}) - 2u(x_i, t_{j+1}) - u(x_{i-1}, t_{j+1})] + O(\Delta t, \Delta x^2).
\] (6.159)

To simplify notation let

\[ r \equiv \frac{\Delta t}{\Delta x^2}. \] (6.160)

The task is to show that the above formulation is unconditionally stable. To achieve this note also that given a matrix \([A]\), eigenvalues \(\lambda_s\) and eigenvectors \(\{v\}_s\) the following relationships hold:

\[
[A] \{v\}_s = \lambda_s \{v\}_s
\] (6.161)

\[
[A]^{-1} \{v\}_s = \frac{1}{\lambda_s} \{v\}_s
\] (6.162)

and that

\[
\lambda_s = 1 + 4r \sin^2 \left( \frac{s\pi}{2N} \right)
\] (6.163)

5. If you were to approximate the equation

\[
\frac{\partial u}{\partial t} + v \frac{\partial u}{\partial x} = 0
\] (6.164)

using a backward in time, backward in space approximation, you require one boundary defined at \(x = 0\) and one initial condition. What would happen if you were to apply a second boundary condition at the terminal end of the interval? What would happen
if you used a central difference approximation for the space derivative and a boundary condition at the terminal end of the interval. To make these determinations, generate a nodal array with one element in the vertical (time) dimension and three in the horizontal (space) dimension and the computational molecule for both formulations and use this information in your explanation.

6. Approximate the equation

\[
\frac{\partial u}{\partial t} + a \frac{\partial^2 u}{\partial x^2} = 0
\]  

(6.165)

using three nodes and a centered finite difference approximation in space and a backward finite difference equation in time for the central spatial node. Now do the same problem using a linear finite element in space and backward difference in time formulation for the central spatial node. What is the fundamental difference in the algebraic equations and describe what you think this means in terms of how the time derivative is handled.

7. Write the finite difference approximation at the location \( x = i\Delta x \) and \( t = \Delta t / 2 \) for the equation

\[
\frac{\partial u}{\partial t} + a \frac{\partial^2 u}{\partial x^2} = 0.
\]

(6.166)

Hint: you will need the space approximation at both the \( t = 0 \) and \( t = \Delta t \) levels.
Bibliography


Chapter 7

Finite Difference Methods in Two Space

In this chapter we will consider the finite difference approximation to a steady-state (time-independent) problem a two-space dimensional system. The “type” equation we will initially consider is the Poisson’s equation

\[
\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = Q(x, y).
\]  

(7.1)

Since this equation is second order in both \(x\) and \(y\), we need two boundary conditions for each coordinate direction. Thus we have, for example,

\[
\begin{align*}
    u(x, y) &= u_0(x, y) & x, y \in \partial \Omega_x \\
    \frac{\partial}{\partial x} u(x, y) &= \frac{\partial}{\partial x} u_1(x, y) & x, y \in \partial \Omega_y \\
    \partial \Omega_x + \partial \Omega_y &= \partial \Omega
\end{align*}
\]

(7.2) \hspace{1cm} (7.3) \hspace{1cm} (7.4)

where \(u_0(x, y)\) and \(\frac{\partial}{\partial x} u_1(x, y)\) are known functions and the notation \(\partial \Omega_x\) and \(\partial \Omega_y\)

Figure 7.1: Domain of interest regarding the boundary conditions specified in Eq. (7.2) through (7.4).

denote the portion of the perimeter as indicated in Fig.7.1.
For the case of rectangular elements and a finite difference approximation, the two-space dimensional case is analogous to the two-dimensional space-time case (see Eq. (6.10)). One simply replaces the time coordinate with a second space coordinate. The approximating function over three nodes becomes

\[ u(x, y) = \sum_{i=-1}^{1} \sum_{j=-1}^{1} u_{i,j} \ell_{i}^{n}(x) \ell_{j}^{m}(y) + E_{nm} \]  

(7.5)

which is illustrated in Fig. 7.2 for the case when the superscripts \( n \) and \( m \) (the degrees of the Lagrange polynomials in \( x \) and \( y \), respectively) are both 1. Figure 7.3 presents the quadratic basis function defined over one element. Notice that we have hijacked the subscript \( i \). Earlier it was identified in general with a node, but in the case of weighted residuals, also the weighting function. Now it represents the \( x \) coordinate for the node \( i \) only, that is \( x = i \Delta x \) but there is no reference to its use as an index for a weighting function. We will only use \( i \) in this way in this section and redefine it in the following section. Note that along both the \( x \) and \( y \) directions the function is linear, but in the

![Figure 7.2: Diagrammatic representation of the function \( \ell_{i}^{1}(x) \ell_{j}^{1}(y) \).

![Figure 7.3: Quadratic polynomials in two dimensions. The circles are the nodes and the black circles indicate the location for which the polynomial is defined, that is, the location at which it equals unity (modified from [Felippa, 2015]).](image)
diagonal directions it is a quadratic because of the product $\ell_i^1(x)\ell_j^1(y)$. We can use these functions to represent the first derivative, but for the second derivative we need to use at least a quadratic-second degree Lagrange polynomial. If we differentiate Eq. (7.5) twice when $n = m = 2$ and evaluate the result at point $x_i = i\Delta x$ and $y_j = j\Delta y$ we obtain

$$
\frac{\partial^2 u}{\partial x^2} \bigg|_{(x_i, y_j)} = \frac{1}{\Delta x^2} [u_{i-1,j} - 2u_{i,j} + u_{i+1,j}] + \frac{\partial^2}{\partial x^2} E_{i,j}
$$

and

$$
\frac{\partial^2 u}{\partial y^2} \bigg|_{(x_i, y_j)} = \frac{1}{\Delta y^2} [u_{i,j-1} - 2u_{i,j} + u_{i,j+1}] + \frac{\partial^2}{\partial y^2} E_{i,j}.
$$

Substitution of Eqs. (7.6) and (7.7) into Eq. (7.1) yields our two-dimensional finite difference approximation of Eq. (7.1) which is given by

$$
\frac{1}{\Delta x^2} [u_{i-1,j} - 2u_{i,j} + u_{i+1,j}] + \frac{1}{\Delta y^2} [u_{i,j-1} - 2u_{i,j} + u_{i,j+1}]
$$

$$
+ \frac{\partial^2}{\partial y^2} E_{i,j} + \frac{\partial^2}{\partial x^2} E_{i,j} = Q(x_i, y_j).
$$

If we neglect the error $E_{i,j}$ this equation becomes

$$
\frac{1}{\Delta x^2} [u_{i-1,j} - 2u_{i,j} + u_{i+1,j}] + \frac{1}{\Delta y^2} [u_{i,j-1} - 2u_{i,j} + u_{i,j+1}] = Q(x_i, y_j)
$$

and the computational molecule for this equation is given in Fig. 7.4. In this figure, the weights for each of the nodal values of $u_{i,u}$ etc. are shown assuming $\Delta x = \Delta y = 1$

![Figure 7.4: Computational molecule for Eq. 7.9 where $\Delta x = \Delta y = 1$.](image)

We will now consider an example to illustrate how this equation is used when a cross-derivative is involved. Consider the generalization of Eq. (7.1). This type of equation arises when the coefficients in the equation are tensors, that is

$$
a_{xx} \frac{\partial^2 u(x,y)}{\partial x^2} + a_{xy} \frac{\partial}{\partial x} \left( \frac{\partial u(x,y)}{\partial y} \right) + a_{yx} \frac{\partial}{\partial y} \left( \frac{\partial u(x,y)}{\partial x} \right) + a_{yy} \frac{\partial^2 u(x,y)}{\partial y^2} = 0.
$$

(7.10)
where coefficients are known constants. Now the finite difference approximation is a little more complicated. We begin by writing the second degree Lagrange approximation for \( \hat{u}(x, y) \) viz.

\[
\hat{u}(x, y) = \sum_{i=0}^{n} \sum_{j=0}^{m} u(x_i, y_j) \ell_i^2(x) \ell_j^2(y).
\]  

(7.11)

Now we differentiate this expression with respect to \( x \) and then \( y \) to get

\[
\frac{\partial \hat{u}(x, y)}{\partial x} = \sum_{i=0}^{n} \sum_{j=0}^{m} u(x_i, y_j) \frac{\partial \ell_i^2(x)}{\partial x} \ell_j^2(y)
\]

(7.12)

\[
\frac{\partial \hat{u}(x, y)}{\partial y} = \sum_{i=0}^{n} \sum_{j=0}^{m} u(x_i, y_j) \ell_i^2(x) \frac{\partial \ell_j^2(y)}{\partial y}.
\]

(7.13)

Next differentiate Eq. (7.12) with respect to \( x \) and Eq. (7.13) with respect to \( y \). Performing these actions we get

\[
\frac{\partial^2 \hat{u}(x, y)}{\partial x^2} = \sum_{i=0}^{n} \sum_{j=0}^{m} u(x_i, y_j) \frac{\partial^2 \ell_i^2(x)}{\partial x^2} \ell_j^2(y)
\]

(7.14)

\[
\frac{\partial^2 \hat{u}(x, y)}{\partial y^2} = \sum_{i=0}^{n} \sum_{j=0}^{m} u(x_i, y_j) \ell_i^2(x) \frac{\partial^2 \ell_j^2(y)}{\partial y^2}.
\]

(7.15)

We can now reverse the procedure and differential Eq. (7.12) with respect to \( y \) and Eq. (7.13) with respect to \( x \). When this is done one obtains

\[
\frac{\partial^2 \hat{u}(x, y)}{\partial y \partial x} = \sum_{i=0}^{n} \sum_{j=0}^{m} u(x_i, t_j) \frac{\partial \ell_i^2(x)}{\partial x} \frac{\partial \ell_j^2(y)}{\partial y}
\]

(7.16)

\[
\frac{\partial^2 \hat{u}(x, y)}{\partial x \partial y} = \sum_{i=0}^{n} \sum_{j=0}^{m} u(x_i, t_j) \frac{\partial \ell_i^2(x)}{\partial x} \frac{\partial \ell_j^2(y)}{\partial y}.
\]

(7.17)

We have evaluated these derivatives above in Eqs. (7.6) and (7.7) and we found

\[
\frac{d\hat{u}^2(x_i, y_j)}{dx^2} = \frac{1}{\Delta x^2} [u_{x_i+1, y_j} - 2u_{x_i, y_j} + u_{x_i-1, y_j}]
\]

(7.18)

\[
\frac{d\hat{u}^2(x_i, y_j)}{dy^2} = \frac{1}{\Delta y^2} [u_{x_{i+1}, y_{j+1}} - 2u_{x_i, y_j} + u_{x_{i-1}, y_{j-1}}].
\]

(7.19)

The cross derivatives are a little tricky. One can show by expanding the derivatives found in Eqs. (7.16) and (7.17) that the complete expansion will consist of nine terms, one for each of the nine nodes found in Fig. [2]. However, all but four of them vanish because of the nature of the nature of the functions. They are zero on at least two sides of the square and, therefore their derivatives are zero along the tangent to those sides. What remains are the terms

\[
\frac{\partial^2 \hat{u}(x, y)}{\partial x \partial y} = \\
\left\{ u(x_{i-1}, y_{j-1}) \frac{\partial \ell_{i-1}^2(x)}{\partial x} \frac{\partial \ell_{j-1}^2(y)}{\partial y} + u(x_{i-1}, y_{j+1}) \frac{\partial \ell_{i-1}^2(x)}{\partial x} \frac{\partial \ell_{j+1}^2(y)}{\partial y} + \\
+ u(x_{i+1}, y_{j-1}) \frac{\partial \ell_{i+1}^2(x)}{\partial x} \frac{\partial \ell_{j-1}^2(y)}{\partial y} + u(x_{i+1}, y_{j+1}) \frac{\partial \ell_{i+1}^2(x)}{\partial x} \frac{\partial \ell_{j+1}^2(y)}{\partial y} \right\}
\]

(7.20)
The various derivatives found in this equation evaluated at \((x_i,y_j)\) are as follows:

\[
\frac{\partial}{\partial x} \ell_{i-1}^2 (x) \bigg|_{x_i} = \left[ \frac{\partial}{\partial x} \left( \frac{(x-x_i)(x-x_{i+1})}{(x_{i-1}-x_i)(x_{i-1}-x_{i+1})} \right) \right] \bigg|_{x_i} = \frac{x_i - x_{i+1}}{(x_{i-1} - x_i)(x_{i-1} - x_{i+1})} = -\frac{\Delta x}{2(\Delta x^2)} = -\frac{1}{2\Delta x} \tag{7.21}
\]

\[
\frac{\partial}{\partial x} \ell_{i+1}^2 (x) \bigg|_{x_i} = \left[ \frac{\partial}{\partial x} \left( \frac{(x-x_{i-1})(x-x_i)}{(x_{i+1}-x_{i-1})(x_{i+1}-x_i)} \right) \right] \bigg|_{x_i} = \frac{x_i - x_{i-1}}{(x_{i+1} - x_{i-1})(x_{i+1} - x_i)} = \frac{\Delta x}{2(\Delta x^2)} = \frac{1}{2\Delta x}. \tag{7.22}
\]

Similarly, for the \(y\) direction we get the following using a similar reasoning

\[
\frac{\partial}{\partial y} \ell_{j-1}^2 (y) \bigg|_{y_j} = -\frac{1}{2\Delta y} \tag{7.23}
\]

\[
\frac{\partial}{\partial y} \ell_{j+1}^2 (y) \bigg|_{y_j} = \frac{1}{2\Delta y}. \tag{7.24}
\]

Substituting the above derivative approximations into Eq. (7.20) we get

\[
\frac{\partial^2 \hat{u}(x,y)}{\partial x \partial y} \bigg|_{x_i,y_j} = u(x_{i-1},y_{j-1}) \left( -\frac{1}{2\Delta x} \right) \left( -\frac{1}{2\Delta y} \right) + u(x_{i-1},y_{j+1}) \left( -\frac{1}{2\Delta x} \right) \left( \frac{1}{2\Delta y} \right) + u(x_{i+1},y_{j-1}) \left( \frac{1}{2\Delta x} \right) \left( -\frac{1}{2\Delta y} \right) + u(x_{i+1},y_{j+1}) \left( \frac{1}{2\Delta x} \right) \left( \frac{1}{2\Delta y} \right) \tag{7.25}
\]

\[
\frac{\partial^2 \hat{u}(x,y)}{\partial x \partial y} \bigg|_{x_i,y_j} = \frac{1}{4\Delta x \Delta y} (u(x_{i-1},y_{j-1})-u(x_{i-1},y_{j+1})+u(x_{i+1},y_{j-1})-u(x_{i+1},y_{j+1})) \tag{7.26}
\]

To get the final form of the approximation we combine Eqs. (7.19), (7.18) and (7.26) to get
\[ \begin{align*}
\frac{a_{xx}}{\Delta x^2} [u_{i+1,j} - 2u_{i,j} + u_{i-1,j}] + \frac{a_{yy}}{\Delta y^2} [u_{i,j+1} - 2u_{i,j} + u_{i,j-1}] + \\
\frac{a_{xy}}{4\Delta x\Delta y} [u(x_{i-1}, y_{j-1}) - u(x_{i-1}, y_{j+1}) + u(x_{i+1}, y_{j+1}) - u(x_{i+1}, y_{j-1})] + \\
\frac{a_{yx}}{4\Delta y\Delta x} [u(x_{i+1}, y_{j+1}) - u(x_{i+1}, y_{j-1}) + u(x_{i-1}, y_{j-1}) - u(x_{i-1}, y_{j+1})]
\end{align*} \] (7.27)

\[ = 0. \] (7.28)

If we assume \( \Delta x = \Delta y = a_{xx} = a_{xy} = a_{yx} = a_{yy} = 1 \) the computational molecule for this cross-derivative bearing equations is given in Fig. 7.5. Notice that now there are nine, not five, nodes in the molecule. The cross derivative added four. From a computational point of view this is not a good thing. The matrix used to solve a problem with this formulation will have four extra bands parallel to the diagonal and this means significantly more effort is needed to solve the resulting set of equations when standard matrix algebra methods are used.

![Computational molecule](image)

Figure 7.5: Computational molecule for Eq. 7.28 when we assume \( \Delta x = \Delta y = a_{xx} = a_{xy} = a_{yx} = a_{yy} = 1 \).

### 7.1 Example Problem

Consider the problem defined by equation

\[ \frac{\partial^2}{\partial x^2} u(x,y) + \frac{\partial^2}{\partial y^2} u(x,y) = 1 \] (7.29)

shown in Fig. 7.6.

Assume boundary conditions

\[ \begin{align*}
u (0, y) &= 1 \quad \text{(7.30)} \\
u (2, y) &= 0 \quad \text{(7.31)} \\
u (x, 0) &= 0 \quad \text{(7.32)} \\
u (x, 2) &= 0. \quad \text{(7.33)}
\end{align*} \]
7.2. Chapter Summary

The finite difference expression for this problem is obtained directly from Eq. (7.8) by assigning the appropriate nodal identifiers. Because only the node \((1,1)\) is unknown in this problem (the rest are known from boundary conditions), the difference equation is

\[
\frac{1}{\Delta x^2} [u_{0,1} - 2u_{1,1} + u_{2,1}] + \frac{1}{\Delta y^2} [u_{1,0} - 2u_{1,1} + u_{1,2}] = 1.
\] (7.34)

Substitution of values known from boundary conditions yields

\[
\frac{1}{l^2} [1 - 2u_{1,1} + 0] + \frac{1}{l^2} [0 - 2u_{1,1} + 0] = 1
\] (7.35)

which simplifies to

\[-4u_{1,1} = 0\] (7.36)

and finally

\[u_{1,1} = 0.\] (7.37)

If we assume there is no flux, that is, \(Q = 0\) we get

\[
\frac{1}{l^2} [1 - 2u_{1,1} + 0] + \frac{1}{l^2} [0 - 2u_{1,1} + 0] = 0
\] (7.38)

which reduces to

\[-4u_{1,1} = -1\] (7.39)

and

\[u_{1,1} = 1/4.\] (7.40)

**7.2 Chapter Summary**

In this brief chapter, we introduce the concept of representing a two-dimensional problem using finite difference approximations and present an elementary example.
7.3 Problems

1. Consider the finite difference mesh indicated in Fig. 7.7. The equation of interest is

\[ \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0 \]  

(7.41)

with boundary conditions

\[ u(0, y) = 1 \]  

(7.42)

\[ \frac{\partial u(4, y)}{\partial x} = 2 \]  

(7.43)

\[ \frac{\partial u(x, 0)}{\partial y} = 0 \]  

(7.44)

\[ \frac{\partial u(x, 4)}{\partial y} = 0 \]  

(7.45)

Do the following:

(a) Write the finite difference expression for Eq. (7.41). You should have five terms in your approximation.

(b) Write a finite difference approximation for the Neumann boundary conditions (Eqs. (7.43)–(7.45)).

(c) Write the finite difference equation for Eq. (7.42). You now have enough equations.

(d) Substitute the information from steps 1b and 1c into step 1a.

(e) Solve the resulting equation for the unknown nodes.

2. The approximation of \( u \) for the case of the two-dimensional element shown in Fig 7.7 is written as

\[ u(x, y) \approx \hat{u}(x, y) = \sum_{i=0}^{2} \sum_{j=0}^{2} u_{i,j} \ell_i(x) \ell_j(y) \]  

(7.46)

where the \( \ell_i(x) \) and \( \ell_j(y) \) are quadratic Lagrange polynomials defined as

\[ \ell_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} \]  

(7.47)

\[ \ell_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} \]  

(7.48)

\[ \ell_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} \]  

(7.49)

with a similar set of definitions for \( \ell_j(y) \), \( i = 0, 1, 2 \).

3. You are to obtain the finite difference representation of

\[ \frac{\partial^2 \hat{u}}{\partial x \partial y}. \]  

(7.50)

Please proceed in the following steps:
Figure 7.7: Mesh and nodal arrangement for problems 1 (panel a) and 2 (panel b).
(a) Differentiate Eq. (7.46) first with respect to $y$ and then the result with respect to $x$. You should have nine terms of the general form

$$u_{i,j} \frac{\partial \ell_i(x)}{\partial x} \frac{\partial \ell_j(y)}{\partial y}.$$  \hspace{1cm} (7.51)

(b) Place these into Eq. (7.46) after differentiation and evaluate the nine terms using Eq. (7.47)

(c) Replace the various spatial increments with the values from the above figure, that is, $x_1 - x_0 = 1$ etc.; be careful with your signs.

(d) Now evaluate the result at the center node $u(1,1)$, since that is where you are writing the approximation

(e) The result should be

$$\frac{(u_{2,2} - u_{0,2} - u_{2,0} + u_{0,0})}{4} = (7.52)$$ \hspace{1cm} (7.52)

$$\frac{(u_{2,2} - u_{0,2} - u_{2,0} + u_{0,0})}{4}.$$ \hspace{1cm} (7.53)
Bibliography


Saulev, V.K. (1957) "On a Class of Elliptic Equations Solvable by the Method of Finite Differences, Vycisl. Mat., 1, 81–86

Chapter 8

Finite Element Methods in Two Space

8.1 Finite Element Approximations over Rectangles

In this section we extend the one-dimensional finite element method of approximation to two space dimensions. We begin with rectangular elements because our earlier theory lends itself most readily to this approach. We begin by once again considering the Poisson’s equation on a rectangular domain

\[
\frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} = Q(x, y)
\]  

\(8.1\)

\[
u(x, y) = u_0(x, y) \quad x, y \in \partial \Omega_x\]  

\(8.2\)

\[
u(x, y) = u_1(x, y) \quad x, y \in \partial \Omega_y\]  

\(8.3\)

\[
\partial \Omega_x + \partial \Omega_y = \partial \Omega.
\]  

\(8.4\)

![Figure 8.1: Definition sketch of the region of interest for Eq. 8.1.](image)

We now digress briefly to introduce some notation. As we have done earlier, let us begin with the approximation of \( u(x, y) \) by \( \hat{u}(x, y) \). We will use the same point of
departure as we used for the finite difference approximations in Chapter 7, that is

$$\hat{u}(x, y) = \sum_{i=0}^{n} \sum_{j=0}^{m} u_{i,j} \ell_{i}^{m}(x) \ell_{j}^{m}(y)$$  \hspace{1cm} (8.5)$$

where we assume we are working with Lagrange polynomials of degree $n$ and $m$. But now we make a change in notation which, while not critical for our current discussion of rectangular elements, will be essential when we talk about triangular elements in the next section. The strategy is to replace the $i, j$ notation for a nodal location in two-dimensional problems with a single index.

Earlier, in the case of problems defined in one space dimension, there was no ambiguity; the $i, j$ index played the role of both nodal identifier and nodal location through the relationship $x = i\Delta x$ and $y = j\Delta y$. However, this is all about to change. To make a long story short, we will replace the double $i, j$ subscript notation for a node location (that is based on the idea of $i$ associated with a row and $j$ associated with a column) with a new $p_{\alpha}$ index where $\alpha$ in this instance could take on the value $i$ or $j$. The $i, j$ notation was convenient because it identified the node by the value $i, j$ and also identified its location $x = i\Delta x$ and $y = j\Delta y$ (assuming $\Delta x$ and $\Delta y$ are constant everywhere.

Note that $p_{\alpha}$ does not, in and of itself, tell you where the node $p_{\alpha}$ is located. We need additional information. For example, we need to know that node number $p_{\alpha}$ is located at $x_{p_{\alpha}}, y_{p_{\alpha}}$. In other words, we need a table that relates our node number $p_{\alpha}$ to its location in terms of $x_{\alpha}$ and $y_{\alpha}$. The index $p_{\alpha}$ is simply a nodal number and, in general, $x \neq p_{\alpha}\Delta x$ and $y \neq p_{\alpha}\Delta y$.

To solidify this idea, consider the node located at $p_{5}$ in Fig. 8.2. Unless we know that $p_{5}$ is located at location $(x_{5}, y_{5}) = (1, 1)$, the knowledge that $p_{\alpha} = p_{5}$ is meaningless. Whereas, using $i, j$ notation we would know both that the node is identified by the unique values of $i = 1$ and $j = 1$, that is, if we have $\ell_{i}^{5}(x) \ell_{j}^{5}(y)$ we know it is located at $(1, 1)$. Somehow using the single index does not seem like progress, but let us go a little further.

Let us take a look at how we would describe a Lagrangian polynomial using this notation. First of all, the Lagrangian polynomial now needs only one subscript to be identified uniquely. Because any two-dimensional Lagrangian is the product of a one-dimensional Lagrangian in $x$ and another in $y$, we can write

$$\ell_{i}(x, y) = \ell_{i}(x) \ell_{i}(y).$$  \hspace{1cm} (8.6)$$

This holds as long as the index $i$ refers to a node, irrespective of where that node is located.

To make the value of Eq. (8.6) unique, we need to know where the node $i$ is located. To illustrate our point, assume function $\ell_{i}(x, y)$ is identified with a node located at $p_{5} = (x_{5}, y_{5})$. Then we know from our earlier discussion that the Lagrange polynomial is identified with a node located at $(x_{5}, y_{5}) = (1, 1)$. If we know this information and the mesh spacing, we have a unique representation of $\ell_{5}(x, y)$ provided we have similar information about all the nodes in the neighborhood (because all connecting node locations are part of the definition of $\ell_{5}(x, y)$). For the case in point, those are all the nodes appearing in Fig. 8.2.

In summary, if you have a Lagrange polynomial (or its derivatives) with subscript $i$, you need to know where node $i$ is located. A table will tell you that node $i$ is located at location $p_{i}$ and that location $p_{i}$ is $x_{p_{i}}, y_{p_{i}}$. You will see that you will need to have
8.1. FINITE ELEMENT APPROXIMATIONS OVER RECTANGLES

this location information when you need to evaluate $\ell_i(x, y)$ when, for example, you need to calculate an integral to obtain a coefficient.

It is helpful at this point to compare Fig. 8.2 with Fig. 7.6 to see the correspondence between the two notations. Employing the double index notation, the approximating function $\tilde{u}(x, y)$ shown in Fig. 7.6 is

$$
\tilde{u}(x, y) = \sum_{i=1}^{3} \sum_{j=1}^{3} u_{ij} \ell_i(x) \ell_j(y) 
$$

(8.7)

and for the grid shown in Fig. 8.2, which uses only one index we have

$$
\tilde{u}(x, y) = \sum_{k=1}^{9} u_k \ell_k(x, y).
$$

(8.8)

We have used the index $k$ in Eq. (8.8) to avoid confusion with the use of the index $i$ in Eq. (8.7). These two expressions, Eqs. (8.7) and (8.8), represent the same information in the special case of rectangular elements, provided the relationship between the index $k$ and the location $p_k$ is provided. So, in the future, when you see an index $i$ or $j$ you must identify each with a unique location $p_i$ or $p_j$ in order to proceed. Note that both indices $i$ and $j$ can take on the same $p$ value since either $i$ or $j$ (or both) may be identified with location $(x_{p_a}, y_{p_a})$.

![Figure 8.2: Single digit numbering for the finite element method of solution.](image)

With the notation issue behind us, let us proceed by stating the residual, which is obtained by replacing $u(x, y)$ with $\tilde{u}(x, y)$ in Eq. (8.1), that is,

$$
\frac{\partial^2}{\partial x^2} \tilde{u}(x, y) + \frac{\partial^2}{\partial y^2} \tilde{u}(x, y) - Q(x, y) = R(x, y).
$$

(8.9)

The appropriate weighted residual formulation is

$$
\int_{\Omega} R(x, y) \ell_i(x, y) d\Omega = 0 \quad i = 1, 2, ..., N
$$

(8.10)
where \( N \) is the number of nodes. Now, substitute for \( R(x, y) \) from Eq. (8.9).

\[
\int_{\Omega} \left[ \frac{\partial^2}{\partial x^2} \hat{u}(x, y) + \frac{\partial^2}{\partial y^2} \hat{u}(x, y) - Q(x, y) \right] \ell_i(x, y) = 0 \quad i = 1, 2, \ldots, N
\]  

(8.11)

Next employ Green’s Theorem (two-dimensional version of integration by parts) to the second-order terms. We get

\[
\int_{\Omega} - \left[ \frac{\partial}{\partial x} \hat{u}(x, y) \frac{\partial}{\partial x} \ell_i(x, y) + \frac{\partial}{\partial y} \hat{u}(x, y) \frac{\partial}{\partial y} \ell_i(x, y) + Q(x, y) \ell_i(x, y) \right] d\Omega
\]  

\[+
\int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \ell_i(x, y) d\ell = 0 \quad i = 1, 2, \ldots, N
\]  

(8.12)

where \( d\ell \) is taken along the side \( \partial \Omega \) in a counterclockwise direction. We will explain how to handle the term

\[
\int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \ell_i(x, y) d\ell
\]

later on page 191. Substitution for \( \hat{u}(x, y) \) yields

\[
\int_{\Omega} - \left[ \sum_{j=1}^{N} u_j \left\{ \frac{\partial}{\partial x} \ell_j(x, y) \frac{\partial}{\partial x} \ell_i(x, y) + \frac{\partial}{\partial y} \ell_j(x, y) \frac{\partial}{\partial y} \ell_i(x, y) \right\} \right.
\]

\[+Q(x, y) \ell_i(x, y) d\Omega \]  

\[+ \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \ell_i(x, y) d\ell = 0 \quad i = 1, 2, \ldots, N.
\]  

(8.13)

Keep in mind that we are now using single valued indices so the subscripts \( i \) and \( j \) are now referring to basis functions and weighting functions located at nodes \( i \) and \( j \).

Let us assume we are working with the same problem presented earlier and reproduced as Fig. 8.2. We now write Eq. (8.13) for this problem

\[
\int_{\Omega} - \left[ \sum_{j=1}^{9} u_j \left\{ \frac{\partial}{\partial x} \ell_j(x, y) \frac{\partial}{\partial x} \ell_i(x, y) + \frac{\partial}{\partial y} \ell_j(x, y) \frac{\partial}{\partial y} \ell_i(x, y) \right\} + Q(x, y) \ell_i(x, y) \right] d\Omega
\]

\[+ \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \ell_i(x, y) d\ell = 0 \quad i = 1, 2, \ldots, 9.
\]  

(8.14)

The matrix form of this equation is

\[
\begin{bmatrix}
  a_{11} & a_{12} & 0 & a_{14} & a_{15} & 0 & 0 & 0 & 0 \\
  a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} & 0 & 0 & 0 \\
  0 & a_{32} & a_{33} & 0 & a_{35} & a_{36} & 0 & 0 & 0 \\
  a_{41} & a_{42} & 0 & a_{44} & a_{45} & 0 & a_{47} & a_{48} & 0 \\
  a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} & a_{57} & a_{58} & a_{59} \\
  0 & a_{62} & a_{63} & 0 & a_{65} & a_{66} & 0 & a_{68} & a_{69} \\
  0 & 0 & 0 & a_{74} & a_{75} & 0 & a_{77} & a_{78} & 0 \\
  0 & 0 & 0 & a_{84} & a_{85} & a_{86} & a_{87} & a_{88} & a_{89} \\
  0 & 0 & 0 & 0 & a_{95} & a_{96} & 0 & a_{98} & a_{99}
\end{bmatrix}
\begin{bmatrix}
  u_1 \\
  u_2 \\
  u_3 \\
  u_4 \\
  u_5 \\
  u_6 \\
  u_7 \\
  u_8 \\
  u_9
\end{bmatrix}
= \begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3 \\
  f_4 \\
  f_5 \\
  f_6 \\
  f_7 \\
  f_8 \\
  f_9
\end{bmatrix}
\]  

(8.15)
8.1. FINITE ELEMENT APPROXIMATIONS OVER RECTANGLES

or in more simplified notation.

\[ [A] \{u\} = \{f\}. \]  \hspace{1cm} (8.16)

The elements of the coefficient matrix \([A]\) are defined as follows

\[ a_{ij} = - \int_{\Omega} \left( \frac{\partial}{\partial x} \ell_j(x, y) \frac{\partial}{\partial x} \ell_i(x, y) + \frac{\partial}{\partial y} \ell_j(x, y) \frac{\partial}{\partial y} \ell_i(x, y) \right) d\Omega \]  \hspace{1cm} (8.17)

\[ f_i = \int_{\Omega} Q(x, y) \ell_i(x, y) d\Omega - \int_{\partial \Omega} \frac{\partial \hat{u}}{\partial n} \ell_i(x, y) d\ell. \]  \hspace{1cm} (8.18)

The information appearing on the right-hand side will be known from a specification of \(Q(x, y)\) or, as will be seen later, by boundary conditions. Now consider the evaluation of the integral appearing in Eq. (8.17). In our earlier work (see Section 5.2 on page 94) we performed the integrations over the elements. We will do the same thing here. We rewrite Eq. (8.17) as

\[ a_{ij} = - \sum_{e=1}^{E} \int_{\Omega_e} \left( \frac{\partial}{\partial x} \ell_j(x, y) \frac{\partial}{\partial x} \ell_i(x, y) + \frac{\partial}{\partial y} \ell_j(x, y) \frac{\partial}{\partial y} \ell_i(x, y) \right) d\Omega_e \]  \hspace{1cm} (8.19)

where the integration is now over element \(\Omega_e\).

Consider a typical element such as shown in Fig. 8.3; in our case it happens to be a unit square, but that is not necessary. Note that element numbers are shown circled. The contribution of this element for node \((i, j)\) is

\[ - \int_{\Omega_1} \left( \frac{\partial}{\partial x} \ell_j(x, y) \frac{\partial}{\partial x} \ell_i(x, y) + \frac{\partial}{\partial y} \ell_j(x, y) \frac{\partial}{\partial y} \ell_i(x, y) \right) d\Omega_1 \]  \hspace{1cm} (8.20)\[ i = 1, 2, 5, 4, \]  \hspace{0.5cm} \( j = 1, 2, 5, 4. \)

We can expand the definitions of the basis and weighting functions using Eq. (8.6) to give

\[ - \int_{\Omega_1} \left[ \frac{\partial}{\partial x} (\ell_j(x) \ell_j(y)) \frac{\partial}{\partial x} (\ell_i(x) \ell_i(y)) + \frac{\partial}{\partial y} (\ell_j(y) \ell_j(x)) \frac{\partial}{\partial y} (\ell_i(y) \ell_i(x)) \right] d\Omega_1 \]  \hspace{1cm} (8.21)\[ i = 1, 2, 5, 4, \]  \hspace{0.5cm} \( j = 1, 2, 5, 4. \)

This looks like a backwards step, but, as we will see in a moment, it actually simplifies our development.

To set the stage for what follows, we will change direction and consider the case of a generic element with node numbering 1, 2, 5, 4 shown in Fig. 8.3. There are a lot of numbers in this figure, so let us take a minute to clarify them. The ones and zeros around the outside of the array of nodes are the boundary condition values. The numbers to the lower right of the nodes are the node numbers; they go from 1 to 9. The numbers in the circles are the element numbers. There are four of them and the one in the lower left of the figure is shaded. At the far left and bottom of the figure are the distances along \(y\) and \(x\), respectively. The ranges are \(x \in [0, 2]\) and \(y = [0, 2]\). Then the translation in numbering between our actual or ‘global’ element and our generic or ‘local element’ is given in the following table. Take a moment to study this little table. In the first row are the nodes defined in the local coordinate system \((\xi, \eta)\). In the second row are the corresponding values in the global coordinate system. Thus, for example, if you look in the second column you see that the local node 1 corresponds to the global node 1 but in the case of column 4 the local node 3 corresponds to the global node 5.
The integrations we need to perform to employ the Galerkin finite element method can be accomplished for this element using local coordinates \( \chi \) for \( x \) and \( \eta \) for \( y \) (see Section 1.8 on page 14 for an introduction to local coordinates):

\[
\begin{array}{cccc}
\text{Local} & 1 & 2 & 3 & 4 \\
\text{Global} & 1 & 2 & 5 & 4 \\
\end{array}
\]

(8.22)

Figure 8.3: Finite difference grid of two-dimensional problem in space.

Figure 8.4: One local element defined in \( \xi, \eta \) local coordinates.
8.1. Finite Element Approximations over Rectangles

\[ - \int_0^1 \int_0^1 \left( \frac{\partial}{\partial x} [\ell_J(\chi) \ell_J(\eta)] \frac{\partial \chi}{\partial x} \right) \left( \frac{\partial}{\partial x} [\ell_I(\chi) \ell_I(\eta)] \frac{\partial \chi}{\partial x} \right) \frac{dx}{d\chi} \frac{dy}{d\eta} d\eta d\chi \\
- \int_0^1 \int_0^1 \left( \frac{\partial}{\partial \eta} [\ell_J(\eta) \ell_J(\chi)] \frac{\partial \eta}{\partial y} \right) \left( \frac{\partial}{\partial \eta} [\ell_I(\eta) \ell_I(\eta)] \frac{\partial \eta}{\partial y} \right) \frac{dx}{d\chi} \frac{dy}{d\eta} d\eta d\chi \\
J = 1, 2, 3, 4 \quad I = 1, 2, 3, 4. \quad (8.23) \]

Expansion of the derivatives and noting that since \( \chi \) and \( \eta \) are orthogonal so that terms such as \( \frac{\partial}{\partial \eta} [\ell_J(\chi)] = 0 \), yield

\[ - \int_0^1 \int_0^1 \left( \frac{\partial}{\partial x} [\ell_J(\chi) \ell_J(\eta)] \frac{\partial \chi}{\partial x} \right) \left( \frac{\partial}{\partial x} [\ell_I(\chi) \ell_I(\eta)] \frac{\partial \chi}{\partial x} \right) \frac{dx}{d\chi} \frac{dy}{d\eta} d\eta d\chi \\
- \int_0^1 \int_0^1 \left( \frac{\partial}{\partial \eta} [\ell_J(\eta) \ell_J(\chi)] \frac{\partial \eta}{\partial y} \right) \left( \frac{\partial}{\partial \eta} [\ell_I(\eta) \ell_I(\eta)] \frac{\partial \eta}{\partial y} \right) \frac{dx}{d\chi} \frac{dy}{d\eta} d\eta d\chi. \quad (8.24) \]

Notice the subtle difference between Eqs. (8.23) and (8.24).

Consider, as an example, the term associated with \( I = 1, J = 1 \), that is

\[ - \int_0^1 \int_0^1 \left( \frac{\partial}{\partial x} [\ell_1(\chi)] \ell_1(\eta) \frac{\partial \chi}{\partial x} \right) \left( \frac{\partial}{\partial x} [\ell_1(\chi)] \ell_1(\eta) \frac{\partial \chi}{\partial x} \right) \frac{dx}{d\chi} \frac{dy}{d\eta} d\eta d\chi. \quad (8.25) \]

From the definitions of the Lagrange polynomials in local coordinates (see Eq. (5.15) on page 85) we have, after performing the required differentiations,

\[ - \int_0^1 \int_0^1 \left( -\frac{1}{2} \ell_1(\eta) \frac{1}{2} \right) \left( -\frac{1}{2} \ell_1(\eta) \frac{1}{2} \right) \frac{1}{2} d\chi \frac{1}{2} d\eta \\
= \int_0^1 \int_0^1 \left( -\frac{1}{2} \ell_1(\chi) \frac{1}{2} \right) \left( -\frac{1}{2} \ell_1(\chi) \frac{1}{2} \right) \frac{1}{2} d\chi \frac{1}{2} d\eta. \quad (8.26) \]

or, on further substitution

\[ - \int_0^1 \int_0^1 \left( -\frac{1}{2} (1 - \eta) \frac{1}{2} \right) \left( -\frac{1}{2} (1 - \eta) \frac{1}{2} \right) \frac{1}{2} d\chi \frac{1}{2} d\eta \\
- \int_0^1 \int_0^1 \left( -\frac{1}{2} (1 - \chi) \frac{1}{2} \right) \left( -\frac{1}{2} (1 - \chi) \frac{1}{2} \right) \frac{1}{2} d\chi \frac{1}{2} d\eta. \quad (8.27) \]

which reduces to

\[ - \int_0^1 \int_0^1 (1 - \eta) (1 - \eta) d\chi d\eta \\
- \int_0^1 \int_0^1 (1 - \chi) (1 - \chi) d\chi d\eta. \quad (8.28) \]

Evaluation of the integrals gives the \((1,1)\) entry of the local coefficient matrix, that is

\[ \left( -\frac{1}{3} \right) + \left( -\frac{1}{3} \right) = \frac{2}{3}. \quad (8.29) \]
As a second example, consider $I = 1, \; J = 2$, which produces an off-diagonal coefficient. From Eq. (8.24) we get

$$-\int_0^1 \int_0^1 \left( \frac{\partial}{\partial \chi} [\ell_2(\chi)] \ell_2(\eta) \right) \cdot \left( \frac{\partial}{\partial \chi} [\ell_1(\chi)] \ell_1(\eta) \right) \frac{dx}{dx} \frac{dy}{dy} \frac{d\eta}{d\eta}$$

$$-\int_0^1 \int_0^1 \left( \frac{\partial}{\partial \eta} [\ell_2(\eta)] \ell_2(\chi) \right) \cdot \left( \frac{\partial}{\partial \eta} [\ell_1(\eta)] \ell_1(\chi) \right) \frac{dx}{dx} \frac{dy}{dy} \frac{d\eta}{d\eta}.$$  \hspace{1cm} (8.30)

Again we substitute for the definitions of the Lagrange polynomials and differentiate to give

$$-\int_0^1 \int_0^1 \left( \frac{1}{1} \ell_2(\eta) \frac{1}{1} \right) \left( -\frac{1}{1} \ell_1(\eta) \frac{1}{1} \right) \frac{1}{1} \frac{dx}{dx} \frac{1}{1} \frac{dy}{dy} \frac{1}{1} \frac{d\eta}{d\eta}$$

$$-\int_0^1 \int_0^1 \left( \frac{1}{1} \ell_2(\chi) \frac{1}{1} \right) \left( -\frac{1}{1} \ell_1(\chi) \frac{1}{1} \right) \frac{1}{1} \frac{dx}{dx} \frac{1}{1} \frac{dy}{dy} \frac{1}{1} \frac{d\eta}{d\eta}.$$ \hspace{1cm} (8.31)

which upon further substitution yields,

$$-\int_0^1 \int_0^1 \left( \frac{1}{1} (1-\eta) \frac{1}{1} \right) \left( -\frac{1}{1} (1-\eta) \frac{1}{1} \right) \frac{1}{1} \frac{dx}{dx} \frac{1}{1} \frac{dy}{dy} \frac{1}{1} \frac{d\eta}{d\eta}$$

$$-\int_0^1 \int_0^1 \left( -\frac{1}{1} (1-\chi) \frac{1}{1} \right) \left( -\frac{1}{1} (1-\chi) \frac{1}{1} \right) \frac{1}{1} \frac{dx}{dx} \frac{1}{1} \frac{dy}{dy} \frac{1}{1} \frac{d\eta}{d\eta}.$$ \hspace{1cm} (8.32)

which simplifies to

$$\int_0^1 \int_0^1 (1-\eta) (1-\eta) d\chi d\eta$$

$$-\int_0^1 \int_0^1 (1-\chi) (1-\chi) d\chi d\eta.$$ \hspace{1cm} (8.33)

and finally yields, upon integration,

$$\left( \frac{1}{3} \right) + \left( -\frac{1}{6} \right) = \frac{1}{6}.$$ \hspace{1cm} (8.34)

If we perform similar integrations for the remaining nodal combinations in this element we obtain the local coefficient matrix which we provide as a table. For convenience in this table, we have identified the nodes in row 2 and also in column 5.

<table>
<thead>
<tr>
<th>J</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-\frac{1}{3}</td>
<td>\frac{1}{3}</td>
<td>\frac{1}{3}</td>
<td>\frac{1}{3}</td>
</tr>
<tr>
<td>2</td>
<td>\frac{1}{3}</td>
<td>-\frac{1}{3}</td>
<td>\frac{1}{3}</td>
<td>\frac{1}{3}</td>
</tr>
<tr>
<td>3</td>
<td>\frac{1}{3}</td>
<td>\frac{1}{3}</td>
<td>-\frac{1}{3}</td>
<td>\frac{1}{3}</td>
</tr>
<tr>
<td>4</td>
<td>\frac{1}{3}</td>
<td>\frac{1}{3}</td>
<td>\frac{1}{3}</td>
<td>-\frac{1}{3}</td>
</tr>
</tbody>
</table>

The form of the matrix is

$$\begin{bmatrix}
-\frac{2}{3} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{6} & -\frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{6} & -\frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{6} & \frac{1}{3} & -\frac{1}{3}
\end{bmatrix}.$$ \hspace{1cm} (8.36)
8.1. **FINITE ELEMENT APPROXIMATIONS OVER RECTANGLES**

All the element matrices will be the same in this case so we can construct the global matrix using the local–global coefficient transformation table, that is the table shown as Eq. (8.37). To be sure we understand this table, let us take a minute to examine it. The second row contains the nodal locations expressed in terms of global coordinates (see Fig. 8.5). The last column on the right contains the element numbers. The third row contains the local coordinate node numbers in element 1 corresponding to the same nodes defined in the global numbering system. For example, global node 5 shown in the second row corresponds to the local node 3 identified in the third row with the local coordinate system for element 1. Since this mapping from global node numbers to local node numbers is a basic concept, the reader is encouraged to review this paragraph until it is clearly understood.

The global matrix is shown in Eq. (8.38). Let us see how it is obtained. Consider the (1, 1) coefficient in Eq. (8.38). This coefficient is associated only with global node 1. We first locate the element in which global node 1 is located (please refer to Fig. 8.5). We see that it is in element number 1. Next find the local node that corresponds to the global node (1, 1) in local element 1; it is local node 1. So the value we seek to fill location (1, 1) in the global matrix is the coefficient (1, 1) in local element matrix 1. Equation (8.35) or Eq. (8.36) tells us that the value we seek is \(-\frac{2}{3}\). Since node 1 in the global matrix appears in only one element, the information in location (1, 1) is all that is available. Thus, placing the value \(-\frac{2}{3}\) in the global matrix at location (1, 1) completes the operation to fill this global coefficient location.

The global location (2, 2) is different because it involves information from both elements 1 and 2. This is because the weighting function for this node spans both elements. Thus, to obtain this coefficient, it is necessary to use the table in Eq. (8.37) for element 1 and 2. For the first element we see that global node 2 corresponds to local node 2. Therefore we need to take the value (2, 2) from the element coefficient matrix and this is again \(-\frac{2}{3}\) and place it in global matrix location (2, 2).

But we are not finished. We now need to obtain the contribution from element 2. Global node 2 in element 2 is local node 1 as seen from the table in row 4, column 3. Going to Eq. (8.36) we again find in element (1, 1) a value of \(-\frac{2}{3}\). We need to add this to the global coefficient location (2, 2). Thus, we have \(-\frac{2}{3}\) from element 1 and \(-\frac{2}{3}\) from element 2 which gives a total of \((-\frac{2}{3} - \frac{2}{3} = -\frac{4}{3}\) which we found in the global matrix at location (2, 2). It would be prudent for the reader to try and obtain the global coefficient (2, 5), which involves yet another concept:

\[
[A] = \begin{bmatrix}
\frac{2}{3} & -\frac{1}{6} & 0 & -\frac{1}{6} & -\frac{1}{3} & 0 & 0 & 0 & 0 \\
-\frac{1}{6} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 \\
0 & -\frac{1}{6} & \frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\
-\frac{1}{6} & \frac{1}{3} & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\
-\frac{2}{3} & -\frac{3}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 \\
0 & -\frac{1}{3} & \frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -\frac{1}{6} & \frac{1}{6} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 \\
\end{bmatrix}
\]  
(8.38)
Figure 8.5: Figure used to relate the location of nodes in the local and global coordinate systems.

Consider now the integral associated with the right-hand side of our matrix equation that includes the source term \( Q(x, y) \), namely

\[
f_i = \int_{\Omega} Q(x, y) \ell_i(x, y) \, d\Omega
\]  

(8.39)

which, assuming \( Q(x, y) \) constant over the element, is for our local element

\[
f_I = Q \int_0^1 \int_0^1 \ell_i(\chi, \eta) \frac{dx}{d\chi} \frac{dy}{d\eta} \, d\eta \, d\chi
\]  

(8.40)

Consider the case of \( I = 1 \)

\[
f_1 = Q \int_0^1 \int_0^1 \ell_1(\chi, \eta) \frac{dx}{d\chi} \frac{dy}{d\eta} \, d\eta \, d\chi
\]  

(8.41)

or, substituting for \( \ell_1(x, y) \), we obtain

\[
f_1 = Q \int_0^1 \int_0^1 \ell_1(\chi) \ell_1(\eta) \frac{dx}{d\chi} \frac{dy}{d\eta} \, d\eta \, d\chi
\]  

(8.42)

Upon substitution of the definitions of the Lagrange polynomials we get

\[
f_1 = Q \int_0^1 \int_0^1 (1 - \chi)(1 - \eta) \frac{dx}{d\chi} \frac{dy}{d\eta} \, d\eta \, d\chi
\]  

(8.43)
8.1. **FINITE ELEMENT APPROXIMATIONS OVER RECTANGLES**

or

\[ f_1 = Q \left[ \int_0^1 (1 - \chi) \frac{dx}{d\chi} d\chi \right] \left[ \int_0^1 (1 - \eta) \frac{dy}{d\eta} d\eta \right] . \]  

Integration of this equation provides our final result, namely

\[ f_1 = Q \left( \frac{1}{2} \right) \left( \frac{1}{2} \right) = \frac{Q}{4} . \]  

In general we also have a term of the form (see Eq. (8.18)) which represents a Neumann, second type or flux boundary condition, that is,

\[ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_i (x, y) d\ell . \]  

Writing this integral in local coordinates we obtain for integration along the \( \eta \) direction

\[ \frac{\partial \tilde{u}}{\partial n} \int_0^1 \ell_1 (\chi, \eta) \frac{dy}{d\eta} d\eta \]  

or

\[ \frac{\partial \tilde{u}}{\partial n} \int_0^1 \ell_1 (\chi) \ell_1 (\eta) \frac{dy}{d\eta} d\eta . \]  

Let us select \( I = 1 \) to provide a concrete example

\[ \frac{\partial \tilde{u}}{\partial n} \ell_1 (\chi) \int_0^1 \ell_1 (\eta) \frac{dy}{d\eta} d\eta = \frac{\partial \tilde{u}}{\partial n} \ell_1 (\chi) \int_0^1 (1 - \eta) \frac{1}{d\eta} \frac{dy}{d\eta} d\eta = \frac{\partial \tilde{u}}{\partial n} \ell_1 (\chi) \left[ \eta - \frac{\eta^2}{2} \right]_0^1 = \frac{\partial \tilde{u}}{\partial n} \left( \frac{1}{2} \right) . \]

Note that the value of \( \ell_1 (\chi) \) at \( \chi = 0 \) is unity and thus the (1) in the last equation. Eq. (8.49) states that one half of the flux entering the side of the element of length 1 is allocated to node 1. The other half would be allocated to the adjacent node along the external face of the element being considered. The \( f(\chi, \eta) \) vector is obtained by examining all the nodes in the global system, that is

\[ \{f\} = \begin{bmatrix} \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_1 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_2 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_3 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_4 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_5 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_6 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_7 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_8 (x, y) d\ell \\ \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \ell_9 (x, y) d\ell \end{bmatrix} . \]
If we combine the information provided above about our approximation of Eq. (8.1), we obtain the matrix equation

$$
\begin{bmatrix}
\frac{2}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 \\
\frac{4}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\
0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\
-\frac{1}{6} & -\frac{1}{3} & 0 & \frac{4}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 \\
-\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & \frac{4}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3}
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
u_4 \\
u_5 \\
u_6 \\
u_7 \\
u_8 \\
u_9
\end{bmatrix}
- \begin{bmatrix}
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_1 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_2 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_3 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_4 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_5 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_6 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_7 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_8 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_9 (x, y) d\ell 
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
1 \\
u_5 \\
0 \\
1 \\
0 \\
0 \\
0
\end{bmatrix}.
$$

Application of the boundary conditions gives

$$
\begin{bmatrix}
\frac{2}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 & 0 \\
\frac{4}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 & 0 \\
0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\
-\frac{1}{6} & -\frac{1}{3} & 0 & \frac{4}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} & 0 \\
-\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & \frac{4}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \\
0 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3}
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
u_4 \\
u_5 \\
u_6 \\
u_7 \\
u_8 \\
u_9
\end{bmatrix}
- \begin{bmatrix}
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_1 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_2 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_3 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_4 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_5 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_6 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_7 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_8 (x, y) d\ell \\
\int_{\partial \Omega} n \frac{\partial u}{\partial n} \ell_9 (x, y) d\ell 
\end{bmatrix}
= \begin{bmatrix}
1 \\
0 \\
0 \\
1 \\
u_5 \\
0 \\
1 \\
0 \\
0 \\
0
\end{bmatrix}.
$$

which simplifies for this simple case to

$$-8u_5 = -1 - 1 + 3Q/1$$

where the surface integral associated with this equation (Eq. (8.53)) is zero because \( \ell_5 (x, y) \) is zero along the external boundary.
8.1. FINITE ELEMENT APPROXIMATIONS OVER RECTANGLES

Assuming $Q = 1$

$$u_5 = 0. \quad (8.54)$$

This is the same result we obtained with the finite difference method.

If we assume $Q = 0$ we have

$$u_5 = \frac{3}{8} \quad (8.55)$$

which compares with $u_5 = \frac{1}{4}$ obtained using the finite difference method. The secret as to why these solutions are different lies in the way that the source term $Q$ is handled. In essence we use a point value at the center node for the finite difference representation and an averaged value in the finite element formulation. This will be better understood after reading the following discussion on how the finite difference and finite element methods differ for this problem.

To see how the finite difference and finite element approximations are different we expand the equation for $u_5$ to obtain

$$\frac{1}{3} (u_1 + u_2 + u_3 + u_4 - 8u_5 + u_6 + u_7 + u_8) = Q \Delta x \Delta y. \quad (8.56)$$

We now modify this equation through addition and extraction of these eight values to give something that resembles a series of second derivative approximations, that is,

$$\frac{1}{6} (u_1 - 2u_2 + u_3) + 4 \frac{u_4 - 2u_5 + u_6}{\Delta x^2} + \frac{u_7 - 2u_8 + u_9}{\Delta x^2}$$

$$+ \frac{1}{6} (u_1 - 2u_4 + u_7) + 4 \frac{u_2 - 2u_5 + u_8}{\Delta y^2} - \frac{u_3 - 2u_6 + u_9}{\Delta y^2} = -Q \Delta x \Delta y. \quad (8.57)$$

Next we multiply and divide the first term by $\Delta x^2$ and the second by $\Delta y^2$. The result is

$$\frac{\Delta x^2}{6} \left[ \frac{(u_1 - 2u_2 + u_3)}{\Delta x^2} + 4 \frac{(u_4 - 2u_5 + u_6)}{\Delta x^2} + \frac{(u_7 - 2u_8 + u_9)}{\Delta x^2} \right]$$

$$+ \frac{\Delta y^2}{6} \left[ \frac{(u_1 - 2u_4 + u_7)}{\Delta y^2} + 4 \frac{(u_2 - 2u_5 + u_8)}{\Delta y^2} - \frac{(u_3 - 2u_6 + u_9)}{\Delta y^2} \right]$$

$$= -Q \Delta x \Delta y \quad (8.58)$$

The next step is to divide the equation through by $\Delta x \Delta y$ to yield

$$\frac{1}{\Delta x \Delta y} \frac{\Delta x^2}{6} \left[ \frac{(u_1 - 2u_2 + u_3)}{\Delta x^2} + 4 \frac{(u_4 - 2u_5 + u_6)}{\Delta x^2} + \frac{(u_7 - 2u_8 + u_9)}{\Delta x^2} \right]$$

$$+ \frac{1}{\Delta x \Delta y} \frac{\Delta y^2}{6} \left[ \frac{(u_1 - 2u_4 + u_7)}{\Delta y^2} + 4 \frac{(u_2 - 2u_5 + u_8)}{\Delta y^2} - \frac{(u_3 - 2u_6 + u_9)}{\Delta y^2} \right]$$

$$= -Q. \quad (8.59)$$
We now simplify and rearrange the terms multiplying the contents of the square brackets. The result is

\[
\frac{1}{2\Delta y} \left[ \frac{(u_1 - 2u_2 + u_3) - 4(u_4 - 2u_5 + u_6) + (u_7 - 2u_8 + u_9)}{\Delta x^2} \right] + \frac{1}{2\Delta x} \left[ \frac{(u_4 - 2u_5 + u_6) + (u_7 - 2u_8 + u_9)}{\Delta y^2} \right] + \frac{1}{2\Delta y} \left[ \frac{(u_1 - 2u_4 + u_7) + 4(u_2 - 2u_5 + u_8) - (u_3 - 2u_6 + u_9)}{\Delta y^2} \right] = -Q. \tag{8.60}
\]

Finally, we recognize that the multipliers of the second derivative approximations are those used in Simpson's rule. We take advantage of this relationship to give the following equation:

\[
\frac{1}{2\Delta y} \left[ \frac{(u_1 - 2u_2 + u_3) - 4(u_4 - 2u_5 + u_6) + (u_7 - 2u_8 + u_9)}{\Delta x^2} \right] + \frac{1}{2\Delta x} \left[ \frac{(u_4 - 2u_5 + u_6) + (u_7 - 2u_8 + u_9)}{\Delta y^2} \right] + \frac{1}{2\Delta y} \left[ \frac{(u_1 - 2u_4 + u_7) + 4(u_2 - 2u_5 + u_8) - (u_3 - 2u_6 + u_9)}{\Delta y^2} \right] + Q \\
\approx \frac{1}{2\Delta y} \int_0^{2\Delta y} \frac{\partial^2 u}{\partial x^2} dy + \frac{1}{2\Delta x} \int_0^{2\Delta x} \frac{\partial^2 u}{\partial y^2} dx + Q. \tag{8.61}
\]

Equation 8.61 states that the finite element formulation for the problem analyzed produces a discrete equation that uses the finite difference approximations to the second derivative terms defined in one coordinate direction averaged over the coordinate direction orthogonal to that of the derivatives.

The template for the finite element algorithm described above is found in Fig. 8.6. It is made up of the template for the \(x\)-derivative (Fig. 8.7) added to the template of the \(y\)-derivative (Fig. 8.8).

![Figure 8.6: Finite element mesh with weights.](image)
8.2 Finite Element Approximations over Triangles

Consider once again the equation

\[ \frac{\partial^2}{\partial x^2} u(x, y) + \frac{\partial^2}{\partial y^2} u(x, y) = Q(x, y) \quad x, y \in \Omega \]  

(8.62)

defined on the irregular domain illustrated in Fig. 8.9. As earlier, the interior of the domain is defined as \( \Omega \) and the perimeter as \( \partial \Omega \)

\[ u(x, y) = u_0(x, y) \quad x, y \in \partial \Omega_1 \]  

(8.63)

\[ u(x, y) = u_1(x, y) \quad x, y \in \partial \Omega_2 \]  

(8.64)

\[ \partial \Omega_1 + \partial \Omega_2 = \partial \Omega. \]  

(8.65)
We begin the analysis by defining the approximating function applicable to this problem, that is

\[ \hat{u}(x,y) = \sum_{j=1}^{N} u_j \phi_j(x,y). \]  

(8.66)

In this expression we have replaced our Lagrangian basis functions with a new functional form \( \phi_j(x,y) \). This is a polynomial basis function defined on a triangular subspace which we denote as a triangular finite element. The triangular element discretization of the domain shown in Fig. 8.9 is shown in Fig. 8.10.

The triangular basis function for a typical node is shown in Fig. 8.11. Note that, as usual, the basis function is unity at the node for which it is defined and zero at all other nodes. In Fig. 8.12 we show the set of three basis functions for the three nodes in the triangle. Note that everywhere they sum to unity. It is important that the elements be numbered in the counterclockwise order for the formulae to be presented below for integration to provide correct answers.

### 8.2.1 Formulation of Triangular Basis Functions

Consider one basis function, for example the one located at node \( I \) in Fig. 8.11. Note that this triangle employs the local indices \( I, I+1, \) and \( I+2 \). Thus the index \( I \) can take on any of the nodal values in global coordinates shown in Fig. 8.10. To be more specific, if we consider the shaded element in Fig. 8.10 we could choose \( I \) to be node number 1, 3, or 10. If we chose \( I = 1 \), then \( I + 1 \) would be 3 and \( I + 2 \) would be 10.

In developing the equations for the triangular basis functions we will require, as noted above, that they be unity at the node for which they are defined and zero at the other two nodes in the triangle. In other words, we have for the basis function at node
8.2. FINE ELEMENT APPROXIMATIONS OVER TRIANGLES

Figure 8.10: Array of finite elements which discretize the area $\Omega$. Note that the boundary of $\Omega$, denoted as $\partial\Omega$, is subdivided into linear elements $\partial\Omega_e$ that are also the sides of the interior triangles. Observe that nodes are located at the vertices of each element.

\[ \phi_I(x_I, y_I) = 1 \]  \hspace{1cm} (8.67)
\[ \phi_I(x_{I+1}, y_{I+1}) = 0 \]  \hspace{1cm} (8.68)
\[ \phi_I(x_{I+2}, y_{I+2}) = 0 \]  \hspace{1cm} (8.69)

where the notation $\phi_I(x_I, y_I)$ denotes the value of $\phi_I$ at the nodal location $(x_I, y_I)$.

To begin our development, note that the equation of a plane defined in the $x-y$ coordinate system is

\[ \phi_I(x, y) = ax + by + c. \]  \hspace{1cm} (8.70)

If we now evaluate this expression at each node, we have

\[ \phi_I(x_I, y_I) = ax_I + by_I + c = 1 \]  \hspace{1cm} (8.71)
\[ \phi_I(x_{I+1}, y_{I+1}) = ax_{I+1} + by_{I+1} + c = 0 \]  \hspace{1cm} (8.72)
\[ \phi_I(x_{I+2}, y_{I+2}) = ax_{I+2} + by_{I+2} + c = 0 \]  \hspace{1cm} (8.73)

which can be written in matrix form as

\[
\begin{bmatrix}
  x_I & y_I & 1 \\
  x_{I+1} & y_{I+1} & 1 \\
  x_{I+2} & y_{I+2} & 1 \\
\end{bmatrix}
\begin{bmatrix}
  a \\
  b \\
  c \\
\end{bmatrix}
= 
\begin{bmatrix}
  1 \\
  0 \\
  0 \\
\end{bmatrix}
\]

or, written symbolically,

\[ [P] \{q\} = \{g\} . \]  \hspace{1cm} (8.75)

The solution to this equation gives the expression of $\phi_I(x, y)$:

\[ \phi_I(x, y) = \frac{x(y_{I+1} - y_{I+2}) + y(x_{I+2} - x_{I+1}) + (x_{I+1}y_{I+2} - x_{I+2}y_{I+1})}{\det [P]} \]  \hspace{1cm} (x, y) \in [\Omega_e] \]  \hspace{1cm} (8.76)
where \( \det [P] \) is the determinant of the matrix \([P]\). A similar formulation leads to equations for \( \phi_{I+1}(x,y) \) and \( \phi_{I+2}(x,y) \), that is

\[
\phi_{I+1}(x,y) = \frac{x(yI+2-yI) + y(xI-xI+2) + (xI+2yI-xIyI+2)}{\det [P]} \quad (x, y) \in [\Omega_e]
\]

and

\[
\phi_{I+2}(x,y) = \frac{x(yI-yI+1) + y(xI+1-xI) + (xIyI+1-xI+1yI)}{\det [P]} \quad (x, y) \in [\Omega_e].
\]

It turns out that there is a simple integration formula for triangular elements as defined above; the formula is

\[
\int_{\Omega_e} \phi_I^{m_I} \phi_{I+1}^{m_{I+1}} \phi_{I+2}^{m_{I+2}} d\Omega_e = 2A \frac{m_I!m_{I+1}!m_{I+2}!}{(m_I + m_{I+1} + m_{I+2} + 2)!} \quad (8.79)
\]

where \( m_I \) is the power to which the \( \phi_I \) polynomial is raised. For example, in the integral

\[
\int_{\Omega_e} \phi_I \phi_{I+1} \phi_{I+2} d\Omega_e \quad (8.80)
\]

\( m_I, m_{I+1}, \) and \( m_{I+2} \) are all unity, while in the integral

\[
\int_{\Omega_e} \phi_I \phi_I \phi_{I+2} d\Omega_e \quad (8.81)
\]

or, written slightly differently,
8.2. Finite Element Approximations Over Triangles

Figure 8.12: Diagrammatic representation of three triangular basis functions, one for each node of the triangle. Note that the values of the triangles sum to unity anywhere on the triangle.

\[ \int_{\Omega_e} \phi_I^2 \phi_{I+1}^0 \phi_{I+2}^1 d\Omega_e \]  

\( m_I, m_{I+1}, \text{ and } m_{I+2} \) are 2, 0, 1, respectively.

The area of the triangle \( A \) is given by

\[ A = \frac{1}{2} \text{det} [P]. \]  

Integration Example

Consider the following integral:

\[ \text{Integral} = \int_{\Omega_e=\Omega_e} \phi_I(x, y) d\Omega_e. \]  

Since the function \( \phi_I \) appears alone in the integral, \( m_I = 1, m_{I+1} = 0, \text{ and } m_{I+2} = 0. \) Using Eq. (8.79) and recalling that the factorial of zero is defined to be 1, we obtain

\[ \int_{\Omega_e=\Omega_e} \phi_I(x, y) d\Omega_e \]

\[ = 2A_e \frac{1 \cdot 1 \cdot 1}{(1 + 0 + 0 + 2)!} \]

\[ = \frac{2}{6} A_e \]

\[ = \frac{A_e}{3}. \]  

(8.85)
8.2.2 Example Problem of Finite Element Approximation over Triangles

The equation of interest is, once again,

$$\frac{\partial^2}{\partial x^2} u(x, y) + \frac{\partial^2}{\partial y^2} u(x, y) = 1$$  \hspace{1cm} (8.86)

where (see Fig. 8.13)

$$u(x, y) = 1 - 0.763x \hspace{0.5cm} x, y \in \partial \Omega_1$$  \hspace{1cm} (8.87)
$$u(x, y) = 1 - 0.38x \hspace{0.5cm} x, y \in \partial \Omega_2$$  \hspace{1cm} (8.88)
$$u(x, y) = 0 \hspace{1cm} x, y \in \partial \Omega_3.$$  \hspace{1cm} (8.89)

These boundary conditions were selected to simplify the calculations by having the area of each triangle be unity and while letting the Dirichlet boundary values vary linearly along the sides of the triangle.

Figure 8.13: Illustration of triangle used as example in this subsection.

We begin by selecting our approximation for $u(x, y)$ as

$$\hat{u}(x, y) = \sum_{j=1}^{4} u_j \phi_j(x, y)$$  \hspace{1cm} (8.90)

where the basis functions $\phi_j(x, y)$ are defined using Eqs. (8.76)–(8.78). As earlier, we now define the residual by substituting the approximation $\hat{u}(x, y)$ into the governing equation, in this case Eq. (8.86) and obtain the residual $R(x, y)$

$$\frac{\partial^2}{\partial x^2} \hat{u}(x, y) + \frac{\partial^2}{\partial y^2} \hat{u}(x, y) - 1 = R(x, y).$$  \hspace{1cm} (8.91)

Using Galerkin’s method we weight the residual by the same set of functions we are using for the basis functions, that is, $\phi_i(x, y)$ $i = 1, ..., 4$, and integrate each of the four products over $\Omega = \Omega_1 + \Omega_2 + \Omega_3$. The result is

$$\int_{\Omega} R(x, y) \phi_i(x, y) d\Omega = 0 \hspace{0.5cm} i = 1, ..., 4.$$  \hspace{1cm} (8.92)
8.2. Finite Element Approximations over Triangles

Substitution of Eqs. (8.90) and (8.91) into (8.92) yields

\[
\int_{\Omega} \left\{ \sum_{j=1}^{4} u_j \left( \frac{\partial^2 \phi_j(x,y)}{\partial x^2} + \frac{\partial^2 \phi_j(x,y)}{\partial y^2} \right) - 1 \right\} \phi_i(x,y) \, d\Omega = 0 \quad i = 1, \ldots, 4. \quad (8.93)
\]

We now apply Green’s theorem to Eq. (8.93) to obtain:

\[
- \sum_{j=1}^{4} u_j \int_{\Omega} \left( \frac{\partial}{\partial x} \phi_j(x,y) \frac{\partial}{\partial x} \phi_i(x,y) + \frac{\partial}{\partial y} \phi_j(x,y) \frac{\partial}{\partial y} \phi_i(x,y) \right) \, d\Omega \\
= \int_{\Omega} (1) \phi_i(x,y) \, d\Omega - \int_{\partial \Omega} \frac{\partial \hat{u}}{\partial n} \phi_i(x,y) \, dl \quad i = 1, \ldots, 4. \quad (8.94)
\]

The equations represented by Eq. (8.93) are of the form

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} \\
  a_{21} & a_{22} & a_{23} & a_{24} \\
  a_{31} & a_{32} & a_{33} & a_{34} \\
  a_{41} & a_{42} & a_{43} & a_{44}
\end{bmatrix}
\begin{bmatrix}
  u_1 \\
  u_2 \\
  u_3 \\
  u_4
\end{bmatrix}
= \begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3 \\
  f_4
\end{bmatrix} \quad (8.95)
\]

or

\[
[A] \{u\} = \{f\} \quad (8.96)
\]

where the coefficients in Eq. (8.95) are defined using Eq. (8.94) as

\[
a_{ij} = - \sum_{e=1}^{3} \int_{\Omega_e} \left( \frac{\partial \phi_j(x,y)}{\partial x} \frac{\partial \phi_i(x,y)}{\partial x} + \frac{\partial \phi_j(x,y)}{\partial y} \frac{\partial \phi_i(x,y)}{\partial y} \right) \, d\Omega_e \quad (8.97)
\]

\[
f_i = - \sum_{e=1}^{3} \int_{\partial \Omega_e} \frac{\partial \hat{u}}{\partial n} \phi_i(x,y) \, dl + \sum_{e=1}^{3} \int_{\Omega_e} (1) \phi_i(x,y) \, d\Omega_e. \quad (8.98)
\]

Note that \( \Omega \) is now represented explicitly using the individual elements denoted by the index \( e \). The basis functions are identified with the element over which the integration is being performed. For example when integrating the basis function defined for node 1 over element 3, \( \phi_i(x,y) \) would appear as \( \phi_1(x,y) \) defined over element 3.

As in our earlier example on page 189 we need a transformation table. Using the \( I \) local index notation introduced for triangles we have

<table>
<thead>
<tr>
<th>Global</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>1</td>
<td>1+1</td>
<td>1+2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>1+1</td>
<td>1</td>
<td>1+2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1+1</td>
<td>1+2</td>
<td>3</td>
</tr>
</tbody>
</table>

Again take a moment to review this table. The global node numbers are provided in the second row and go from 1 to 4. From the third row we find that the \( I^{th} \) node for triangle 1 was selected as the global node 1. Once this decision is made, the other two local nodes, namely \( I+1 \) and \( I+2 \), are automatically assigned to global node numbers 2 and 3, respectively. We know this order to be the case not only because of the table, but also because we know that we must number the nodes in each element in counterclockwise order.
Element Coefficient Matrix

Because we are using triangles, as indicated in the above table, there will be three nodes for each element and therefore the element coefficient matrix will have nine entries, that is

$$
\begin{bmatrix}
    a_{I,I} & a_{I,I+1} & a_{I,I+2} \\
    a_{I+1,I} & a_{I+1,I+1} & a_{I+1,I+2} \\
    a_{I+2,I} & a_{I+2,I+1} & a_{I+2,I+2}
\end{bmatrix}.
$$

(8.99)

The corresponding coefficient matrix for element 1, using global numbering would be

$$
A_1 = \begin{bmatrix}
    a_{11} & a_{12} & a_{13} \\
    a_{21} & a_{22} & a_{23} \\
    a_{31} & a_{32} & a_{33}
\end{bmatrix}
$$

(8.100)

as can also be seen from the above table. The integral associated with coefficient $a_{11}$ in the matrix defined by Eq. (8.100) is obtained from Eq. (8.97) using the local I numbering system:

$$
- \int_{\Omega_1} \left( \frac{\partial}{\partial x} \phi_I (x,y) \frac{\partial}{\partial x} \phi_I (x,y) + \frac{\partial}{\partial y} \phi_I (x,y) \frac{\partial}{\partial y} \phi_I (x,y) \right) d\Omega.
$$

(8.101)

The next step is to use Eq. (8.76), reproduced below, to obtain the derivatives needed:

$$
\phi_I (x,y) = \frac{x(y_{I+1} - y_{I+2}) + y(x_{I+2} - x_{I+1}) + (x_{I+1}y_{I+2} - x_{I+2}y_{I+1})}{\det [P]}, \quad x, y \in [\Omega_e]
$$

(8.102)

from which we obtain upon differentiation

$$
\frac{\partial}{\partial x} \phi_I (x,y) = \frac{(y_{I+1} - y_{I+2})}{\det [P]}
$$

(8.103)

and

$$
\frac{\partial}{\partial y} \phi_I (x,y) = \frac{(x_{I+2} - x_{I+1})}{\det [P]}
$$

(8.104)

where

$$
[P] = \begin{bmatrix}
    x_I & y_I & 1 \\
    x_{I+1} & y_{I+1} & 1 \\
    x_{I+2} & y_{I+2} & 1
\end{bmatrix}.
$$

(8.105)

We now substitute Eqs. (8.103) and (8.104) into (8.101) to get

$$
- \int_{\Omega_1} \left( \frac{\partial}{\partial x} \phi_I (x,y) \frac{\partial}{\partial x} \phi_I (x,y) + \frac{\partial}{\partial y} \phi_I (x,y) \frac{\partial}{\partial y} \phi_I (x,y) \right) d\Omega
= \frac{-1}{(\det [P])^2} \int_{\Omega_1} ((y_{I+1} - y_{I+2})(y_{I+1} - y_{I+2}) + (x_{I+2} - x_{I+1})(x_{I+2} - x_{I+1})) d\Omega.
$$

(8.106)

Using our global-local coordinate table, and recalling that

$$
\det [P] = 2A_e
$$

(8.107)
this term becomes
\[-\frac{1}{(\det [P])^2} \int_{\Omega_1} ((y_{I+1} - y_{I+2}) (y_{I+1} - y_{I+2}) + (x_{I+2} - x_{I+1}) (x_{I+2} - x_{I+1})) d\Omega\]
\[= -\frac{1}{4A_e^2} \int_{\Omega_1} ((y_2 - y_3) (y_2 - y_3) + (x_3 - x_2) (x_3 - x_2)) d\Omega\]
\[= -\frac{1}{4A_e^2} ((y_2 - y_3) (y_2 - y_3) + (x_3 - x_2) (x_3 - x_2)) \int_{\Omega_1} d\Omega\]
\[= -\frac{1}{4} ((-0.76) (-0.76) + (1.31) (1.31))
\[= -0.573. \quad (8.108)\]

As a second example of element integration, let us evaluate coefficient $a_{21}$ of element 1. The required integral becomes
\[-\int_{\Omega_1} \left( \frac{\partial}{\partial x} \phi_I (x, y) \frac{\partial}{\partial x} \phi_{I+1} (x, y) + \frac{\partial}{\partial y} \phi_I (x, y) \frac{\partial}{\partial y} \phi_{I+1} (x, y) \right) d\Omega. \quad (8.109)\]

From Eqs. (8.76) and (8.77) on page 198 we have
\[\frac{\partial}{\partial x} \phi_I (x, y) = \frac{(y_{I+1} - y_{I+2})}{\det [P]} \quad (8.110)\]
\[\frac{\partial}{\partial y} \phi_I (x, y) = \frac{(x_{I+2} - x_{I+1})}{\det [P]} \quad (8.111)\]
\[\frac{\partial}{\partial x} \phi_{I+1} (x, y) = \frac{(y_{I+2} - y_I)}{\det [P]} \quad (8.112)\]
\[\frac{\partial}{\partial y} \phi_{I+1} (x, y) = \frac{(x_I - x_{I+2})}{\det [P]} \quad (8.113)\]
The required integral becomes (recalling that the area of each triangle is, for convenience, designed to be unity)
\[-\int_{\Omega_1} \left( \frac{\partial}{\partial x} \phi_I (x, y) \frac{\partial}{\partial x} \phi_{I+1} (x, y) + \frac{\partial}{\partial y} \phi_I (x, y) \frac{\partial}{\partial y} \phi_{I+1} (x, y) \right) d\Omega\]
\[= -\frac{1}{(\det [P])^2} \int_{\Omega_1} (((y_{I+1} - y_{I+2}) (y_{I+2} - y_I) + (x_{I+2} - x_{I+1}) (x_I - x_{I+2})) d\Omega\]
\[= -\frac{1}{(\det [P])^2} (((y_{I+1} - y_{I+2}) (y_{I+2} - y_I) + (x_{I+2} - x_{I+1}) (x_I - x_{I+2})) \int_{\Omega_1} d\Omega\]
\[= -\frac{1}{4A_1} (((y_{I+1} - y_{I+2}) (y_{I+2} - y_I) + (x_{I+2} - x_{I+1}) (x_I - x_{I+2}))
\[= -\frac{1}{4A_1} ((y_2 - y_3) (y_3 - y_1) + (x_3 - x_2) (x_1 - x_3))
\[= -\frac{1}{4} ((-0.76) (-1.51) + (1.31) (0))
\[= -0.286. \quad (8.114)\]
Global Coefficient Matrix

When calculations such as shown above are completed for each of the nodes and each of the elements, the following element coefficient matrices are generated.

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} \\
  a_{21} & a_{22} & a_{23} \\
  a_{31} & a_{32} & a_{33}
\end{bmatrix}
= - \begin{bmatrix}
  0.573 & 0.286 & -0.85 \\
  0.286 & 0.57 & -0.856 \\
  -0.85 & -0.856 & 1.71
\end{bmatrix}
\tag{8.115}
\]

\[
\begin{bmatrix}
  a_{33} & a_{32} & a_{34} \\
  a_{23} & a_{22} & a_{24} \\
  a_{43} & a_{42} & a_{44}
\end{bmatrix}
= - \begin{bmatrix}
  1.73 & -0.868 & -0.868 \\
  -0.868 & 0.573 & 0.286 \\
  -0.868 & 0.286 & 0.573
\end{bmatrix}
\tag{8.116}
\]

\[
\begin{bmatrix}
  a_{11} & a_{13} & a_{14} \\
  a_{31} & a_{33} & a_{34} \\
  a_{41} & a_{43} & a_{44}
\end{bmatrix}
= - \begin{bmatrix}
  0.573 & -0.849 & 0.276 \\
  -0.849 & 1.71 & -0.868 \\
  0.276 & -0.868 & 0.573
\end{bmatrix}
\tag{8.117}
\]

The next step is to assemble the global matrix by taking the coefficients for each node from all the elements and adding them together as illustrated below.

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} \\
  a_{21} & a_{22} & a_{23} & a_{24} \\
  a_{31} & a_{32} & a_{33} & a_{34} \\
  a_{41} & a_{42} & a_{43} & a_{44}
\end{bmatrix}
= - \begin{bmatrix}
  0.573 + 0.573 & 0.286 & -0.859 - 0.849 & 0.276 \\
  0.286 & 0.570 + 0.573 & -0.856 - 0.868 & 0.286 \\
  -0.859 - 0.849 & -0.856 - 0.868 & 1.71 + 1.73 + 1.71 & -0.868 - 0.868. \\
  0.276 & 0.286 & -0.868 - 0.868 & 0.573 + 0.573
\end{bmatrix}
\tag{8.118}
\]

Before we can solve the problem, we need to evaluate the forcing function

\[
\int_\Omega (1) \phi_i (x, y) \, d\Omega - \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \phi_i (x, y) \, dl
= \frac{1}{d} \int_\Omega \phi_i (x, y) \, d\Omega - \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \phi_i (x, y) \, dl
\tag{8.119}
\]

which is the same as

\[
\int_\Omega \phi_i (x, y) \, d\Omega - \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \phi_i (x, y) \, dl
= \sum_{e=1}^{3} \int_{\Omega_e} \phi_i^e (x, y) \, d\Omega - \int_{\partial \Omega} \frac{\partial \tilde{u}}{\partial n} \phi_i (x, y) \, dl.
\tag{8.120}
\]

From Eq. (8.85) we have

\[
\int_{\Omega_e} \phi_i^e (x, y) \, d\Omega = \frac{A_e}{3}
\tag{8.121}
\]
so we can rewrite Eq. (8.120) for each node (again keeping in mind that the area of each element in this problem is unity).

\[
\begin{align*}
    f_1 &= \sum_{e=1}^{3} \int_{\Omega_e} \phi_1^e (x, y) \, d\Omega - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_1 (x, y) \, dl \\
    &= \frac{A_1}{3} + \frac{A_3}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_1 (x, y) \, dl \\
    &= \frac{2}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_1 (x, y) \, dl \\
    &= \frac{2}{3} \left( \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_1 (x, y) \, dl \right) (8.122)
\end{align*}
\]

\[
\begin{align*}
    f_2 &= \sum_{e=1}^{3} \int_{\Omega_e} \phi_2^e (x, y) \, d\Omega - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_2 (x, y) \, dl \\
    &= \frac{A_1}{3} + \frac{A_2}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_2 (x, y) \, dl \\
    &= \frac{2}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_2 (x, y) \, dl \\
    &= \frac{2}{3} \left( \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_2 (x, y) \, dl \right) (8.123)
\end{align*}
\]

\[
\begin{align*}
    f_3 &= \sum_{e=1}^{3} \int_{\Omega_e} \phi_3^e (x, y) \, d\Omega - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_3 (x, y) \, dl \\
    &= \frac{A_1}{3} + \frac{A_2}{3} + \frac{A_3}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_3 (x, y) \, dl \\
    &= \frac{3}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_3 (x, y) \, dl \\
    &= \frac{3}{3} \left( \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_3 (x, y) \, dl \right) (8.124)
\end{align*}
\]

\[
\begin{align*}
    f_4 &= \sum_{e=1}^{3} \int_{\Omega_e} \phi_4^e (x, y) \, d\Omega - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_4 (x, y) \, dl \\
    &= \frac{A_2}{3} + \frac{A_3}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_4 (x, y) \, dl \\
    &= \frac{2}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_4 (x, y) \, dl \\
    &= \frac{2}{3} \left( \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_4 (x, y) \, dl \right) (8.125)
\end{align*}
\]

Thus, the right-hand side vector becomes

\[
\begin{bmatrix}
    f_1 \\
    f_2 \\
    f_3 \\
    f_4
\end{bmatrix} =
\begin{bmatrix}
    \frac{2}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_1 (x, y) \, dl \\
    \frac{3}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_2 (x, y) \, dl \\
    \frac{3}{3} + 0 \\
    \frac{2}{3} - \int_{\partial\Omega} \frac{\partial \hat{u}}{\partial n} \phi_4 (x, y) \, dl
\end{bmatrix} . (8.126)
\]

You may be wondering what we will do with the normal derivative appearing on the right-hand side of Eq. (8.126). As you will see, if we need it, it is provided as a boundary condition. If we do not have a derivative-type boundary condition, this term will not be needed.
Solution of Equations

The matrix equation now becomes

\[
\begin{bmatrix}
1.14 & 0.286 & -1.70 & 0.276 \\
0.286 & 1.14 & -1.72 & 0.286 \\
-1.70 & -1.72 & 5.16 & -1.74 \\
0.276 & 0.286 & -1.74 & 1.15
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
u_4
\end{bmatrix}
= \begin{bmatrix}
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_1 (x, y) \, dl \\
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_2 (x, y) \, dl \\
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_3 (x, y) \, dl \\
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_4 (x, y) \, dl
\end{bmatrix}.
\]

Application of boundary conditions provides the following:

\[
\begin{bmatrix}
1.14 & 0.286 & -1.70 & 0.276 \\
0.286 & 1.14 & -1.72 & 0.286 \\
-1.70 & -1.72 & 5.16 & -1.74 \\
0.276 & 0.286 & -1.74 & 1.15
\end{bmatrix}
\begin{bmatrix}
0 \\
1 \\
u_3 \\
0
\end{bmatrix}
= \begin{bmatrix}
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_1 (x, y) \, dl \\
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_2 (x, y) \, dl \\
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_3 (x, y) \, dl \\
\frac{2}{3} - \int_{\partial\Omega} \frac{\partial^2}{\partial n^2} \phi_4 (x, y) \, dl
\end{bmatrix}.
\]

which can be reduced to

\[-(5.16) u_3 = [(-1.70) (0) - (1.72) (1) - (1.74) (0)] + 1 \quad (8.129)\]

or

\[u_3 = \frac{-0.72}{-5.16} = 0.14. \quad (8.130)\]

Notice that the equations containing the derivative on the right-hand side have been eliminated by virtue of the specification of the function value of \( u(x, y) \). Thus, because we did not have any Neumann boundary conditions, the equations containing this information disappeared by virtue of the alternative specification of the Dirichlet condition. We will see in the next subsection what to do when there is a Neumann condition specified. The complete approximation for \( u(x, y) \) for this problem becomes

\[
\hat{u}(x, y) = \sum_{j=1}^{4} u_j \phi_j (x, y)
= (0) \phi_1 (x, y) + (1) \phi_2 (x, y) + (0.14) \phi_3 (x, y) + (0) \phi_4 (x, y). \quad (8.131)
\]

8.2.3 Second Type or Neumann Boundary-Value Problem

In this section we will modify our earlier problem presented in Section 8.2.2 to consider a Neumann boundary condition. The problem definition is provided below:

\[
\frac{\partial^2}{\partial x^2} u(x, y) + \frac{\partial^2}{\partial y^2} u(x, y) = 1
\quad (8.132)
\]

where (see Fig. 8.13)

\[
\frac{\partial}{\partial n} u(x, y) = 1 \quad x, y \in \partial \Omega_1
\quad (8.133)

u(x, y) = 0 \quad x, y \in \partial \Omega_2
\quad (8.134)

u(x, y) = 0 \quad x, y \in \partial \Omega_3
\quad (8.135)
8.2. FINITE ELEMENT APPROXIMATIONS OVER TRIANGLES

where we have assumed that nodes 1 and 2 are associated with \( \partial \Omega_1 \). The only changes we need to make to our earlier formulation is to modify the right-hand side to accommodate the new boundary conditions. The new boundary specifications are shown in Fig. 8.14. We first write the earlier formulation

\[
\begin{bmatrix}
1.14 & 0.286 & -1.70 & 0.276 \\
0.286 & 1.14 & -1.72 & 0.286 \\
-1.70 & -1.72 & 5.16 & -1.74 \\
0.276 & 0.286 & -1.74 & 1.15
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
u_4
\end{bmatrix}
= \frac{2}{3} \int_{\partial \Omega} \frac{\partial u}{\partial n} \phi_1 (x, y) \, dl
\]

The terms of interest are in the boxes in Eq. (8.136).

Consider the term associated with the first row, that is the equation identified with node 1 and therefore weighting function 1. The first thing to note is that the triangular basis function shown in Fig. 8.11 is a linear Lagrange polynomial along the side of the triangle. Thus the following holds

\[
\int_{\partial \Omega_1} \frac{\partial u}{\partial n} \phi_1 (x, y) \, dl = \int_{\partial \Omega_1} \frac{\partial u}{\partial n} \phi_1 (x, y) \, dl.
\]

If we assume \( \frac{\partial u}{\partial n} \) constant along the boundary segment \( \partial \Omega_1 \), which is the normal procedure in linear triangular elements, it can be taken out of the integral and we obtain

\[
\int_{\partial \Omega_1} \frac{\partial u}{\partial n} \phi_1 (x, y) \, dl = \frac{\partial u}{\partial n} \int_{\partial \Omega_1} \phi_1 (x, y) \, dl.
\]

The function \( \phi_1 (x, y) \), which is defined only along the boundary \( \partial \Omega_1 \) of the triangle, will be a linear Lagrange polynomial. We use the symbol \( \sigma \) to denote distance along this side. In other words, \( \sigma \) is the one-dimensional coordinate defined along the line \( \partial \Omega_1 \). Using this notation we obtain

\[
\frac{\partial u}{\partial n} \int_{\partial \Omega_1} \phi_1 (\sigma) \, dl = \frac{\partial u}{\partial n} \int_{\partial \Omega_1} \frac{\sigma - \sigma_2}{\sigma_1 - \sigma_2} \, dl
\]

where \( \sigma_1 \) and \( \sigma_2 \) are the locations of nodes 1 and 2 in the \( \sigma \) coordinate system. In Fig. 8.11 we see, for example, that \( \sigma_1 \) and \( \sigma_2 \) could correspond to nodes \( I \) and \( I+1 \) and the integration would be performed along the line connecting them. Integration now
The interpretation of Eq. (8.140) is that the portion of the contribution of \( \frac{\partial u}{\partial n} \) associated with node 1 is \( \frac{\partial u}{\partial n} \sigma_2 \). Thus the integral in the first row on the right-hand side of Eq. (8.136) will take on the value presented in the last line of Eq. 8.140. The coordinate \( \sigma_2 \) is determined by inspection from Fig. 8.14 to be 2.63 length units (see the length of the side of the triangle along \( x \)) and \( \frac{\partial u}{\partial n} = 1 \). Thus we have

\[
\frac{\partial u}{\partial n} \sigma_2 = \frac{1}{2} \cdot \frac{2.63}{2} = 1.32.
\]

(8.141)

A similar argument can be used to obtain \( \int_{\partial \Omega} \frac{\partial u}{\partial n} \phi_2 (x, y) \, dl \) and this turns out to be the same value, namely 1.32.

We are now in a position to modify our matrix to accommodate the boundary conditions. We obtain

\[
\begin{bmatrix}
1.14 & 0.286 & -1.70 & 0.276 \\
0.286 & 1.14 & -1.72 & 0.286 \\
-1.70 & -1.72 & 5.16 & -1.74 \\
0.276 & 0.286 & -1.74 & 1.15
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
0
\end{bmatrix}
= \begin{bmatrix}
\frac{2}{3} - 1.32 \\
\frac{2}{3} - 1.32 \\
\frac{2}{3} - 0 \\
\frac{2}{3} - \int_{\partial \Omega} \frac{\partial u}{\partial n} \phi_4 (x, y) \, dl
\end{bmatrix}.
\]

(8.142)

You will notice that the integral in the last row remains unevaluated. We will see in the next step that the reason for this is that it will not be needed. Using the information from the Dirichlet boundary condition that \( u_4 = 0 \), as we implicitly did in the last example Section 8.2.2, we can multiply this value by the last column of the matrix, put the resulting information on the right-hand side of the matrix equation, and eliminate the last row and last column of the matrix. Note that we now have three unknowns since nodes 1 and 2 are not specified \( u(x, y) \) values. Simplifying Eq. (8.142) we have

\[
\begin{bmatrix}
1.14 & 0.286 & -1.70 \\
0.286 & 1.14 & -1.72 \\
-1.70 & -1.72 & 5.16 \\
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
= \begin{bmatrix}
\frac{2}{3} - 1.32 + (0.276) (0) \\
\frac{2}{3} - 1.32 + (0.286) (0) \\
\frac{2}{3} + 0 - (1.74) (0)
\end{bmatrix}.
\]

(8.143)
8.2. FINE ELEMENT APPROXIMATIONS OVER TRIANGLES

Figure 8.14: Repeated for reference, illustration of triangle used as example in this subsection, that is Section 8.2.2.

or

\[
\begin{bmatrix}
0.86 & 1.14 & -1.72 \\
1.40 & 0.286 & -1.70 \\
-1.70 & -1.72 & 5.16
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
=
\begin{bmatrix}
-0.654 \\
-0.654 \\
1
\end{bmatrix}.
\] (8.144)

Taking the inverse of the coefficient matrix, that is

\[
\begin{bmatrix}
0.86 & 1.14 & -1.72 \\
1.40 & 0.286 & -1.70 \\
-1.70 & -1.72 & 5.16
\end{bmatrix}^{-1}
\begin{bmatrix}
2.25 & 1.10 & 1.11 \\
1.10 & 2.30 & 1.13 \\
1.11 & 1.13 & 0.934
\end{bmatrix}
\] (8.145)

we can solve Eq. (8.144) directly as follows:

\[
\begin{bmatrix}
u_1 \\
u_2 \\
u_3
\end{bmatrix}
=
\begin{bmatrix}
2.25 & 1.10 & 1.11 \\
1.10 & 2.30 & 1.13 \\
1.11 & 1.13 & 0.934
\end{bmatrix}
\begin{bmatrix}
-0.654 \\
-0.654 \\
1
\end{bmatrix}
\]

Thus the solution to our problem is \([u_1, u_2, u_3]^T = [1.09, 1.09, 0.53]^T\) where \(T\) denotes, transpose.

To understand the results, one must remember that the vector \(\mathbf{n}\) associated with \(\partial u / \partial n\) is outward directed. Thus, if \(u\) is a state variable, such as mass, there is mass attempting to enter the system along-side \(\partial \Omega_1\). Thus we would expect that the values of \(u\) near \(\partial \Omega_1\) would be higher than further away, say, at node 3. However, we also have a flux \(Q = 1\) exiting the system that must be satisfied, in part through inward flow along the boundary. Node 4 must remain at zero because that is the fixed boundary condition value for \(u\) at that node.

If we assume that \(Q = 0\), that is there is no source term in the equation, we get
\[
\begin{bmatrix}
1.14 & 0.286 & -1.70 \\
0.286 & 1.14 & -1.72 \\
-1.70 & -1.72 & 5.16 \\
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
0 - 1.32 + (0.276) (0) \\
0 - 1.32 + (0.286) (0) \\
0 + 0 - (1.73) (0) \\
\end{bmatrix} \\
(8.147)
\]

or
\[
\begin{bmatrix}
1.14 & 0.29 & -1.70 \\
0.29 & 1.14 & -1.72 \\
-1.70 & -1.72 & 5.16 \\
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
-1.32 \\
-1.32 \\
0 \\
\end{bmatrix} \\
(8.148)
\]

or
\[
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix} = 
\begin{bmatrix}
2.25 & 1.10 & 1.11 \\
1.10 & 2.30 & 1.13 \\
1.11 & 1.13 & 0.934 \\
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix} \\
(8.149)
\]

or
\[
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix} = 
\begin{bmatrix}
4.42 \\
4.48 \\
2.95 \\
\end{bmatrix} \\
(8.150)
\]

One now observes that the slope of the surface has changed. The high side along \(\partial \Omega_1\) is even higher since this is the side from which the mass flows inwards. If the domain were rectangular, we would expect a uniform slope across the domain from the left to the right. However, the domain is triangular and so the surface is not perfectly flat, but it is close.

As a variant on this example, let us see what happens if we again remove the uniform flux, that is let \(Q = 0\), but also reverse the sense of the flux boundary condition, that is set
\[
\frac{\partial}{\partial n} u(x, y) = -1 \quad x, y \in \partial \Omega_1 \\
u(x, y) = 0 \quad x, y \in \partial \Omega_2 \\
u(x, y) = 0 \quad x, y \in \partial \Omega_3. \\
(8.151-8.153)
\]

The matrix equation becomes
\[
\begin{bmatrix}
1.14 & 0.286 & -1.70 \\
0.286 & 1.14 & -1.72 \\
-1.70 & -1.72 & 5.16 \\
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix}
= 
\begin{bmatrix}
0 + 1.32 + (0.276) (0) \\
0 + 1.32 + (0.286) (0) \\
0 - (1.73) (0) \\
\end{bmatrix} \\
(8.154)
\]

for which the solution is
\[
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix} = 
\begin{bmatrix}
2.25 & 1.10 & 1.11 \\
1.10 & 2.30 & 1.13 \\
1.11 & 1.13 & 0.934 \\
\end{bmatrix}
\begin{bmatrix}
u_1 \\
u_2 \\
u_3 \\
\end{bmatrix} \\
= 
\begin{bmatrix}
4.42 \\
4.49 \\
2.96 \\
\end{bmatrix} \\
(8.155)
\]

or \((u_1, u_2, u_3)^T = -(4.42, 4.49, 2.96)^T\). One now observes that the slope of the surface has changed. The low side is along \(\partial \Omega_2\) since this is the side from which the mass flows outward. If the domain were rectangular, we would expect a uniform positive slope across the domain from the left to the right.
8.3 Isoparametric Finite Element Approximation

In this section we will introduce the concept of isoparametric finite elements. These elements are characteristically irregular in their geometry. They are essentially deformed rectangles. However, the concept can be further extended to allow for sides that have curvature. They are useful in some specific engineering applications. A necessary tool for understanding this concept is the idea of a 'natural coordinate system' which is a special form of local coordinate system and which we will now introduce.

8.3.1 Natural Coordinate Systems

One-Dimensional Natural Coordinate Systems

A natural coordinate system is a local coordinate system (see page 16) which is special in that it permits the specification of a point within an element by a set of dimensionless numbers whose magnitudes never exceed unity. Consider the information presented in Fig. 8.15. The variable $x$ is a global coordinate and $L$ is a local coordinate which is also a natural coordinate. Notice that any point along the indicated interval $[x_1, x_2]$ can be specified by a number whose absolute magnitude does not exceed unity in the $L$ coordinate system.

Our first task is to determine how to relate the natural and local coordinate systems one to the other. From Fig. 8.15 we observe the following: a point $x_p$ somewhere within the interval $(x_1, x_2)$ is related to a point $L_p$ located in the interval $(-1, 1)$ by the expression

$$\frac{x_p}{x_2 - x_1} = \frac{L_p}{1 - (-1)} = \frac{L_p}{2}. \quad (8.156)$$

In general, for any $x$, we have

$$\frac{x - x_1}{x_2 - x_1} = \frac{L - (-1)}{2}. \quad (8.157)$$
We can solve for \( x \) in Eq. (8.157) and obtain
\[
x = x_1 + \frac{L + 1}{2} (x_2 - x_1).
\]
(8.158)

A little algebra gives
\[
x = \frac{1}{2} [2x_1 - Lx_1 - x_1 + (L + 1)x_2].
\]
(8.159)

We now rearrange the terms in this expression with malice of forethought to give
\[
x = \frac{1}{2} (1 - L) x_1 + \frac{1}{2} (1 + L) x_2.
\]
(8.160)

Eq. (8.160) can also be written as
\[
x = \sum_{i=1}^{2} x_i \phi_i (L)
\]
(8.161)

where
\[
\phi_1 (L) = \frac{1 - L}{2}
\]
(8.162)
\[
\phi_2 (L) = \frac{1 + L}{2}.
\]
(8.163)

Let us now look more carefully at the function \( \phi_i (L) \). We observe the following:
\[
\phi_1 (L = -1) = 1
\]
(8.164)
\[
\phi_1 (L = 1) = 0
\]
(8.165)
\[
\phi_2 (L = -1) = 0
\]
(8.166)
\[
\phi_2 (L = 1) = 1.
\]
(8.167)

It appears that the values of \( \phi_i (L) \) satisfy the requirements of basis functions for the interpolation of the coordinate \( x \) in terms of the end points of the interval (we did something similar in Section 1.5 on page 4)!

Two-Dimensional Natural Coordinate Systems

The one-dimensional transformation just presented was rather straight-forward. However, the two-dimensional version is a little more complicated. The local and global elements are shown in Figs. 8.16 and 8.17, respectively. Now let us examine the relationship between them. To be a one-to-one transformation we will require the following:

<table>
<thead>
<tr>
<th>Node Number</th>
<th>Local Coordinate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((-1, -1))</td>
</tr>
<tr>
<td>2</td>
<td>((1, -1))</td>
</tr>
<tr>
<td>3</td>
<td>((1, 1))</td>
</tr>
<tr>
<td>4</td>
<td>((-1, 1))</td>
</tr>
</tbody>
</table>

(8.168)

where the symbol \( \rightarrow \) is to be interpreted as ‘corresponds to.’ Building upon what we learned from our one-dimensional example, namely Eq. (8.161), we want to be able to
determine the coordinates of $x$ and $y$ within the global element using a relationship of the form:

\[
x = \sum_{i=1}^{4} x_i \phi_i (\xi, \eta) \\
y = \sum_{i=1}^{4} y_i \phi_i (\xi, \eta).
\]

We will return to this relationship shortly. However, for now let us consider a different strategy for representing $x$.

![Figure 8.16: Local element in local $\xi - \eta$ coordinate system.](image)

Let us impose the constraint that a line that forms an element side that is a straight line in the global coordinate system is also a straight line in the local coordinate system. If this is to be true, we need to be able to have $x$ vary linearly along $\xi$ and $y$ vary linearly along $\eta$. A function of $\xi$ and $\eta$ that varies linearly along $\xi$ will have the form

\[
x = \alpha_1 + \alpha_2 \xi + \alpha_3 \eta + \alpha_4 \xi \eta.
\]

To see that this equation satisfies the requirement of being linear along $\xi$, substitute $\eta = 1$. This gives

\[
x = \alpha_1 + \alpha_2 \xi + \alpha_3 + \alpha_4 \xi.
\]

which is a linear function of $\xi$. Similarly if we substitute $\eta = -1$ we have

\[
x = \alpha_1 + \alpha_2 \xi - \alpha_3 - \alpha_4 \xi.
\]
Both Eqs. (8.172) and (8.173) provide a linear change in \( x \) with a linear change in \( \xi \), thus satisfying the requirement that an element side that is linear in the global system is linear in the local system.

One can, using a similar argument, show that the appropriate form of the interpolator for \( y \) is

\[
y = \beta_1 + \beta_2 \xi + \beta_3 \eta + \beta_4 \xi \eta.
\]  

(8.174)

It can be shown that when \( \xi \) is \( \pm 1 \), \( y \) is a linear function of \( \eta \).

We would like to find out what the values of \( \alpha \) and \( \beta \) are. To achieve this we impose on Eqs. (8.173) and (8.174) the requirement of one-to-one correspondence between the nodal locations in the global and local coordinate systems as shown in the above table. We obtain by substituting the corresponding values of \( x, \xi \), and \( \eta \) into Eq. (8.173) the following matrix equation for the unknown values of \( a_i, i = 1, ..., 4 \):

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
\end{bmatrix} =
\begin{bmatrix}
  1 & \xi_1 & \eta_1 & \xi_1 \eta_1 \\
  1 & \xi_2 & \eta_2 & \xi_2 \eta_2 \\
  1 & \xi_3 & \eta_3 & \xi_3 \eta_3 \\
  1 & \xi_4 & \eta_4 & \xi_4 \eta_4 \\
\end{bmatrix} \begin{bmatrix}
  \alpha_1 \\
  \alpha_2 \\
  \alpha_3 \\
  \alpha_4 \\
\end{bmatrix}
\]  

(8.175)

Substitution of the local coordinate values for \( \xi \) and \( \eta \) yields

\[
\begin{bmatrix}
  x_1 \\
  x_2 \\
  x_3 \\
  x_4 \\
\end{bmatrix} =
\begin{bmatrix}
  1 & -1 & -1 & 1 \\
  1 & 1 & -1 & -1 \\
  1 & 1 & 1 & 1 \\
  1 & -1 & 1 & -1 \\
\end{bmatrix} \begin{bmatrix}
  \alpha_1 \\
  \alpha_2 \\
  \alpha_3 \\
  \alpha_4 \\
\end{bmatrix}.
\]  

(8.176)

Inverting the coefficient matrix provides the solution for \( \{a\}^T \), that is,
Recall from Eq. (8.169) that we can express \( x \) as a function of the nodal coordinates via the expression

\[
\begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4
\end{bmatrix} = \frac{1}{4} \begin{bmatrix}
1 & 1 & 1 & 1 \\
-1 & 1 & 1 & -1 \\
-1 & -1 & 1 & 1 \\
1 & -1 & 1 & -1
\end{bmatrix} \begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix}.
\]  
(8.177)

But, we have another representation of the \( \alpha \) vector from Eq. (8.177), so we can substitute that relationship into Eq. (8.179) to obtain

\[
x = \frac{1}{4} [1, \xi, \eta, \xi \eta] \begin{bmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4
\end{bmatrix}.
\]  
(8.179)

And from Eq. (8.171)

\[
x = [1, \xi, \eta, \xi \eta] \begin{bmatrix}
x_1 \\
x_2 \\
x_3 \\
x_4
\end{bmatrix}
\]  
(8.178)
We now examine Eqs. (8.178) and (8.180) and see that both of these equations describe the behavior of \( x \). Thus we can equate their right-hand-sides to give the relationship

\[
[\phi_1(\xi, \eta), \phi_2(\xi, \eta), \phi_3(\xi, \eta), \phi_4(\xi, \eta)] = \frac{1}{4} [1, \xi, \eta, \xi \eta] \begin{bmatrix}
1 & 1 & 1 & 1 \\
-1 & 1 & 1 & -1 \\
-1 & -1 & 1 & 1 \\
1 & -1 & 1 & -1
\end{bmatrix}. \tag{8.181}
\]

To obtain further insight into this relationship, consider the functions \( \phi_i(\xi, \eta) \). We have from Eq. (8.181)

\[
\begin{align*}
\phi_1(\xi, \eta) &= \frac{1}{4} [(1) (1) - (1) (\xi) - (1) (\eta) + \xi \eta] \\
&= \frac{1}{4} (1 - \xi - \eta + \xi \eta) \\
&= \left( \frac{1 - \xi}{2} \right) \left( \frac{1 - \eta}{2} \right) \tag{8.182}
\end{align*}
\]

\[
\begin{align*}
\phi_2(\xi, \eta) &= \frac{1}{4} (1 + \xi - \eta - \xi \eta) \\
&= \left( \frac{1 + \xi}{2} \right) \left( \frac{1 - \eta}{2} \right) \tag{8.183}
\end{align*}
\]

\[
\begin{align*}
\phi_3(\xi, \eta) &= \frac{1}{4} (1 + \xi + \eta + \xi \eta) \\
&= \left( \frac{1 + \xi}{2} \right) \left( \frac{1 + \eta}{2} \right) \tag{8.184}
\end{align*}
\]

\[
\begin{align*}
\phi_4(\xi, \eta) &= \frac{1}{4} [1 - \xi + \eta - \xi \eta] \\
&= \left( \frac{1 - \xi}{2} \right) \left( \frac{1 + \eta}{2} \right). \tag{8.185}
\end{align*}
\]

The above functions are called Coordinate Transformation Functions.

**Example Problem**

Consider the point located with the X in the lower panel of Fig. 8.18. The goal is to find the global coordinate location for this point. The first step is to write the general form of the global-local transformation. One obtains global coordinates in terms of the global coordinate node locations and the local coordinate transformation functions as

\[
x = \sum_{i=1}^{4} x_i \phi_i(\xi, \eta) \tag{8.186}
\]

\[
y = \sum_{i=1}^{4} y_i \phi_i(\xi, \eta). \tag{8.187}
\]

Next substitute for the \( x_i \) and \( y_i \) values to get

\[
x = 0 [\phi_1(\xi, \eta)] + 2 [\phi_2(\xi, \eta)] + 2 [\phi_3(\xi, \eta)] + 0 [\phi_4(\xi, \eta)] \tag{8.188}
\]
which gives, using Eqs. (8.182)–(8.185)

\[
x_{|\xi=0.25, \eta=0.25} = 2 \left[ \frac{(1 + 0.25)(1 - 0.25)}{4} \right] + 2 \left[ \frac{(1 + 0.25)(1 + 0.25)}{4} \right] = 0.468 + 0.781 = 1.25.
\] (8.189)

Similarly for the \(y_1\) coordinate location

\[
y = 1 [\phi_1 (\xi, \eta)] + 0 [\phi_2 (\xi, \eta)] + 3 [\phi_3 (\xi, \eta)] + 2 [\phi_4 (\xi, \eta)]
\] (8.190)

\[
y_{|\xi=0.25, \eta=0.25} = 1 \left[ \frac{(1 - 0.25)(1 - 0.25)}{4} \right] + 3 \left[ \frac{(1 + 0.25)(1 + 0.25)}{4} \right] + 2 \left[ \frac{(1 - 0.25) \times (1 + 0.25)}{4} \right] = 0.141 + 1.17 + 0.469 = 1.78.
\] (8.191)

Figure 8.18: Element in local and global coordinates used for example problem. The \(\xi\) in the lower panel maps to the \(x\) in the upper panel via Eqs. 8.186 and 8.187.

So, the calculations indicate that the location \((\xi, \eta) = (0.25, 0.25)\) maps to \((x, y) = (1.25, 1.78)\), as shown in the upper panel of Fig. 8.18. Now that we have the coordinate transformation functions, let us consider the basis functions.

### 8.3.2 Basis Functions

As usual, we are seeking to represent our approximating function \(\hat{u}(x, y)\) by a series employing the product of coefficients and basis functions. Consider for example the
approximation of \( u(x, y) \) over one element

\[
\hat{u}(x, y) = \sum_{k=1}^{4} u_k \psi_k(\xi, \eta)
\]

(8.192)

where \( \psi_k(\xi, \eta) \) are the basis functions. To ensure that a function is approximated linearly along a linear side (see Eq. (8.172)) we consider the relationship

\[
\psi_i = \alpha_{1i} + \alpha_{2i} \xi + \alpha_{3i} \eta + \alpha_{4i} \xi \eta \quad i = 1, 2, 3, 4.
\]

(8.193)

Consider as an example the conditions that apply to basis functions in general and \( \psi_1(\xi, \eta) \) in particular, namely that

\[
\psi_1(\xi_1, \eta_1) = 1
\]

(8.194)

and

\[
\psi_1(\xi_p, \eta_p) = 0 \quad p = 2, 3, 4.
\]

(8.195)

Combination of these constraints with the definition provided in Eq. (8.193) yields, for \( \psi_1(\xi, \eta) \),

\[
\begin{bmatrix}
\psi_1(\xi_1, \eta_1) \\
\psi_1(\xi_2, \eta_2) \\
\psi_1(\xi_3, \eta_3) \\
\psi_1(\xi_4, \eta_4)
\end{bmatrix} =
\begin{bmatrix}
1 & \xi_1 & \eta_1 & \xi_1 \eta_1 \\
1 & \xi_2 & \eta_2 & \xi_2 \eta_2 \\
1 & \xi_3 & \eta_3 & \xi_3 \eta_3 \\
1 & \xi_4 & \eta_4 & \xi_4 \eta_4
\end{bmatrix}
\begin{bmatrix}
\alpha_{11} \\
\alpha_{21} \\
\alpha_{31} \\
\alpha_{41}
\end{bmatrix} =
\begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}
\]

(8.196)

or

\[
\begin{bmatrix}
\psi_1(\xi_1, \eta_1) \\
\psi_1(\xi_2, \eta_2) \\
\psi_1(\xi_3, \eta_3) \\
\psi_1(\xi_4, \eta_4)
\end{bmatrix} =
\begin{bmatrix}
1 & -1 & -1 & 1 \\
1 & 1 & -1 & -1 \\
1 & 1 & 1 & 1 \\
1 & -1 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
\alpha_{11} \\
\alpha_{21} \\
\alpha_{31} \\
\alpha_{41}
\end{bmatrix} =
\begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}.
\]

(8.197)

Solving for the \( \alpha \) vector we obtain

\[
\begin{bmatrix}
\alpha_{11} \\
\alpha_{21} \\
\alpha_{31} \\
\alpha_{41}
\end{bmatrix} = \frac{1}{4}
\begin{bmatrix}
1 & 1 & 1 & 1 \\
-1 & 1 & 1 & -1 \\
-1 & -1 & 1 & 1 \\
1 & -1 & 1 & -1
\end{bmatrix}
\begin{bmatrix}
1 \\
0 \\
0 \\
0
\end{bmatrix}
\]

(8.198)

From Eq. (8.193) we have for \( \psi_1(\xi, \eta) \)

\[
\psi_1 = \alpha_{11} + \alpha_{21} \xi + \alpha_{31} \eta + \alpha_{41} \xi \eta
\]

(8.199)

or using Eq. (8.198)

\[
\psi_1 = \frac{1}{4} - \frac{1}{4} \xi - \frac{1}{4} \eta + \frac{1}{4} \xi \eta.
\]

(8.200)
8.3. ISOPARAMETRIC FINITE ELEMENT APPROXIMATION

In order to demonstrate a concept, we now rearrange this equation to give

$$\psi_1 = \frac{1}{4} (1 - \xi - \eta + \xi \eta)$$
$$= \frac{1}{2} (1 - \xi) \frac{1}{2} (1 - \eta). \quad (8.201)$$

Thus we see by comparison of Eq. (8.201) to Eq. (8.182) that $\phi_1 (\xi, \eta)$ and $\psi_1 (\xi, \eta)$ are identical. Thus $\phi_i (\xi, \eta)$ (or $\psi_i (\xi, \eta)$) is both the coordinate transformation function and the basis function.

8.3.3 Calculation of the Jacobian

To see how to use the isoparametric element we will address an example problem. Consider the equation

$$\frac{\partial^2}{\partial x^2} u(x, y) + \frac{\partial^2}{\partial y^2} u(x, y) - \frac{\partial}{\partial x} u(x, y) - \frac{\partial}{\partial y} u(x, y) = 0. \quad (8.202)$$

Using our standard approach we can write an approximation for $u(x, y)$ as

$$\hat{u}(x, y) = \sum_{j=1}^{n} u_j \phi_j (x, y). \quad (8.203)$$

From Galerkin’s method we have

$$\int_{\Omega} \left( \frac{\partial^2}{\partial x^2} \hat{u}(x, y) + \frac{\partial^2}{\partial y^2} \hat{u}(x, y) - \frac{\partial}{\partial x} \hat{u}(x, y) - \frac{\partial}{\partial y} \hat{u}(x, y) \right) \phi_i (x, y) d\Omega = 0 \quad i = 1, \ldots, n. \quad (8.204)$$

Application of Green’s theorem to the second-order terms yields

$$\int_{\Omega} \left( - \frac{\partial}{\partial x} \hat{u}(x, y) \frac{\partial}{\partial x} \phi_i (x, y) - \frac{\partial}{\partial y} \hat{u}(x, y) \frac{\partial}{\partial y} \phi_i (x, y) \right) d\Omega -$$

$$\int_{\Omega} \left( \frac{\partial}{\partial x} \hat{u}(x, y) \phi_i (x, y) + \frac{\partial}{\partial y} \hat{u}(x, y) \phi_i (x, y) \right) d\Omega$$

$$+ \int_{\partial\Omega} \frac{\partial \hat{u}(x, y)}{\partial n} \phi_i (x, y) = 0 \quad i = 1, \ldots, n. \quad (8.205)$$

Substituting Eq. (8.203) into Eq. (8.204), and introducing the local indices $I$ and $J$, we get for each element that consists of four nodes

$$\int_{\Omega_e} \left( \frac{\partial}{\partial x} \phi_J (\xi, \eta) \frac{\partial}{\partial x} \phi_I (\xi, \eta) + \frac{\partial}{\partial y} \phi_J (\xi, \eta) \frac{\partial}{\partial y} \phi_I (\xi, \eta) \right) d\Omega$$

$$\int_{\partial\Omega_e} \frac{\partial \phi_J (\xi, \eta)}{\partial n} \phi_I (\xi, \eta) + \int_{\partial\Omega_e} \frac{\partial \phi_J (\xi, \eta)}{\partial n} \phi_I (\xi, \eta) d\Omega = 0 \quad I = 1, \ldots, 4 \quad (8.207)$$

where $n$ is now the outward-directed normal to the element face. Notice that this is not an equation equal to zero because we are considering only one element and we would need to sum the information contributed by all the elements associated with node $I$ before we could write an equation for node $I$. 

Typical integrals in Eq. (8.207) are of the form

$$\int_{\Omega_e} \left( \frac{\partial}{\partial x} \phi_J (\xi, \eta) \frac{\partial}{\partial x} \phi_I (\xi, \eta) \right) d\Omega$$

(8.208)

and

$$\int_{\Omega_e} \left( \frac{\partial}{\partial x} \phi_J (\xi, \eta) \phi_I (\xi, \eta) \right) d\Omega.$$  

(8.209)

Let us see how we evaluate these integrals. Since

$$\xi = \xi (x, y)$$

(8.210)

and

$$\eta = \eta (x, y)$$

(8.211)

from the chain rule we have

$$\frac{\partial}{\partial \xi} \phi_J (\xi, \eta) = \frac{\partial \phi_J (\xi, \eta)}{\partial x} \frac{\partial x}{\partial \xi} + \frac{\partial \phi_J (\xi, \eta)}{\partial y} \frac{\partial y}{\partial \xi}$$

(8.212)

$$\frac{\partial}{\partial \eta} \phi_J (\xi, \eta) = \frac{\partial \phi_J (\xi, \eta)}{\partial x} \frac{\partial x}{\partial \eta} + \frac{\partial \phi_J (\xi, \eta)}{\partial y} \frac{\partial y}{\partial \eta}.$$  

(8.213)

We can assemble these equations into matrix form to obtain

$$\begin{bmatrix} \frac{\partial \phi_J (\xi, \eta)}{\partial \xi} \\ \frac{\partial \phi_J (\xi, \eta)}{\partial \eta} \end{bmatrix} = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} \begin{bmatrix} \frac{\partial \phi_J (\xi, \eta)}{\partial x} \\ \frac{\partial \phi_J (\xi, \eta)}{\partial y} \end{bmatrix}.$$  

(8.214)

By convention we define the Jacobian matrix $[J]$ as

$$[J] = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix}$$

(8.215)

such that

$$\begin{bmatrix} \frac{\partial \phi_J (\xi, \eta)}{\partial \xi} \\ \frac{\partial \phi_J (\xi, \eta)}{\partial \eta} \end{bmatrix} = [J] \begin{bmatrix} \frac{\partial \phi_J (\xi, \eta)}{\partial x} \\ \frac{\partial \phi_J (\xi, \eta)}{\partial y} \end{bmatrix}.$$  

(8.216)

However we see in Eqs. (8.212) and (8.213) that we need terms of the form

$$\frac{\partial}{\partial x} \phi_J (\xi, \eta)$$

(8.217)

and

$$\frac{\partial}{\partial y} \phi_J (\xi, \eta).$$

(8.218)

To obtain this form, we need to take the inverse of $[J]$ to yield

$$\begin{bmatrix} \frac{\partial \phi_J (\xi, \eta)}{\partial x} \\ \frac{\partial \phi_J (\xi, \eta)}{\partial y} \end{bmatrix} = [J]^{-1} \begin{bmatrix} \frac{\partial \phi_J (\xi, \eta)}{\partial \xi} \\ \frac{\partial \phi_J (\xi, \eta)}{\partial \eta} \end{bmatrix}.$$  

(8.219)

Computing $[J]^{-1}$ we have
\[ [J]^{-1} = \frac{1}{\text{det} [J]} \begin{bmatrix} \frac{\partial y}{\partial \eta} & -\frac{\partial y}{\partial \xi} \\ -\frac{\partial x}{\partial \eta} & \frac{\partial x}{\partial \xi} \end{bmatrix}. \] (8.220)

Combination of Eqs. (8.219) and (8.220) yields

\[ \left[ \frac{\partial \phi_j(\xi,\eta)}{\partial \phi_j(\xi,\eta)} \right] = \frac{1}{\text{det} [J]} \left[ \frac{\partial y}{\partial \xi} \frac{\partial x}{\partial \xi} - \frac{\partial y}{\partial \eta} \frac{\partial x}{\partial \eta} \right] \left[ \frac{\partial \phi_j(\xi,\eta)}{\partial \xi} \frac{\partial \phi_j(\xi,\eta)}{\partial \eta} \right]. \] (8.221)

The next problem is to determine how to get the Jacobian matrix \([J]\)?

Recall that

\[ x = \sum_{I=1}^{4} x_I \phi_I(\xi,\eta) \] (8.222)

and

\[ y = \sum_{I=1}^{4} y_I \phi_I(\xi,\eta). \] (8.223)

We can now use the relationships in Eqs. (8.222) and (8.223) and write

\[ \frac{\partial x}{\partial \xi} = \sum_{I=1}^{4} x_I \frac{\partial}{\partial \xi} \phi_I(\xi,\eta) \]
\[ = x_1 \left[ \frac{1}{4} (-1 + \eta) \right] + x_2 \left[ \frac{1}{4} (1 - \eta) \right] + x_3 \left[ \frac{1}{4} (1 + \eta) \right] + x_4 \left[ \frac{1}{4} (-1 - \eta) \right] \] (8.224)

\[ \frac{\partial x}{\partial \eta} = \sum_{I=1}^{4} x_I \frac{\partial}{\partial \eta} \phi_I(\xi,\eta) \]
\[ = x_1 \left[ \frac{1}{4} (-1 + \xi) \right] + x_2 \left[ \frac{1}{4} (1 - \xi) \right] + x_3 \left[ \frac{1}{4} (1 + \xi) \right] + x_4 \left[ \frac{1}{4} (1 - \xi) \right] \] (8.225)

\[ \frac{\partial y}{\partial \xi} = \sum_{I=1}^{4} y_I \frac{\partial}{\partial \xi} \phi_I(\xi,\eta) \]
\[ = y_1 \left[ \frac{1}{4} (-1 + \eta) \right] + y_2 \left[ \frac{1}{4} (1 - \eta) \right] + y_3 \left[ \frac{1}{4} (1 + \eta) \right] + y_4 \left[ \frac{1}{4} (-1 - \eta) \right] \] (8.226)

\[ \frac{\partial y}{\partial \eta} = \sum_{I=1}^{4} y_I \frac{\partial}{\partial \eta} \phi_I(\xi,\eta) \]
\[ = y_1 \left[ \frac{1}{4} (-1 + \xi) \right] + y_2 \left[ \frac{1}{4} (1 - \xi) \right] + y_3 \left[ \frac{1}{4} (1 + \xi) \right] + y_4 \left[ \frac{1}{4} (1 - \xi) \right] \] (8.227)
\[
\frac{\partial y}{\partial \eta} = \sum_{I=1}^{4} y_I \frac{\partial}{\partial \eta} \phi_I (\xi, \eta) \\
= y_1 \left[ \frac{1}{4} (\xi - 1) \right] + y_2 \left[ \frac{1}{4} (-1 - \xi) \right] + y_3 \left[ \frac{1}{4} (1 + \xi) \right] + y_4 \left[ \frac{1}{4} (1 - \xi) \right].
\] (8.228)

We can now evaluate the Jacobian. We have

\[
\begin{bmatrix}
\frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\
\frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta}
\end{bmatrix} = \frac{1}{4} \begin{bmatrix}
(-1 + \eta) & (1 - \eta) & (1 + \eta) & (-1 - \eta) \\
(\xi - 1) & (-1 - \xi) & (1 + \xi) & (1 - \xi)
\end{bmatrix} \begin{bmatrix}
x_1 & y_1 \\
x_2 & y_2 \\
x_3 & y_3 \\
x_4 & y_4
\end{bmatrix}. \quad (8.229)
\]

One can also show that

\[
d \xi \, d \eta = \det [J] \, d \xi \, d \eta
\] (8.230)

and

\[
\det [J] = \left[ \frac{\partial y}{\partial \eta} \frac{\partial x}{\partial \xi} - \frac{\partial x}{\partial \eta} \frac{\partial y}{\partial \xi} \right]. \quad (8.231)
\]

We are now in a position to evaluate the integrals of interest. Consider as an example

\[
\int_{\Omega_e} \left( \frac{\partial}{\partial x} \phi_J (\xi, \eta) \frac{\partial}{\partial x} \phi_I (\xi, \eta) + \frac{\partial}{\partial y} \phi_J (\xi, \eta) \frac{\partial}{\partial y} \phi_I (\xi, \eta) \right) \, d \xi \, d \eta. \quad (8.232)
\]

We can write the integral in natural coordinates using Eqs. (8.212)–(8.221) and Eqs.(8.230)–(8.231) as

\[
\int_{\Omega_e} \left( \frac{\partial}{\partial \xi} \phi_J (\xi, \eta) \frac{\partial}{\partial \xi} \phi_I (\xi, \eta) + \frac{\partial}{\partial \eta} \phi_J (\xi, \eta) \frac{\partial}{\partial \eta} \phi_I (\xi, \eta) \right) \, d \xi \, d \eta \\
= \int_{-1}^{1} \int_{-1}^{1} \frac{1}{\det [J]} \left( \frac{\partial}{\partial \xi} \phi_J (\xi, \eta) \frac{\partial y}{\partial \eta} - \frac{\partial x}{\partial \eta} \phi_J (\xi, \eta) \frac{\partial y}{\partial \xi} \right) \\
+ \frac{1}{\det [J]} \left( - \frac{\partial}{\partial \xi} \phi_J (\xi, \eta) \frac{\partial x}{\partial \eta} + \frac{\partial x}{\partial \eta} \phi_J (\xi, \eta) \frac{\partial x}{\partial \xi} \right) \\
+ \frac{1}{\det [J]} \left( - \frac{\partial}{\partial \eta} \phi_I (\xi, \eta) \frac{\partial x}{\partial \eta} + \frac{\partial x}{\partial \eta} \phi_I (\xi, \eta) \frac{\partial x}{\partial \xi} \right) \det [J] \, d \xi \, d \eta. \quad (8.233)
\]

The challenge now is to evaluate the integrals appearing on the right-hand side of Eq. (8.233). This is accomplished using Gauss–Legendre quadrature introduced in Section 3.1.6 on page 59.
8.3.4 Example of Isoparametric Formulation

Let us consider the problem of solving for the temperature distribution in the $x-y$ plane. The problem is defined in Fig. 8.19. The governing equation is

$$K \left( \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) = 0$$ \hfill (8.234)

with boundary conditions

$$T(x, y) = 6 \quad x = 0 \quad y \in [0, 5]$$ \hfill (8.235)
$$T(x, y) = 12 \quad x = 20 \quad y \in [0, 5]$$ \hfill (8.236)
$$\frac{\partial T}{\partial n} = 0 \quad y = 0 \quad x \in (0, 20)$$ \hfill (8.237)
$$\frac{\partial T}{\partial n} = 0 \quad y = 5 \quad x \in (0, 20).$$ \hfill (8.238)

**Galerkin Formulation**

Let us assume the temperature can be represented by the series

$$\hat{T}(x, y) = \sum_{j=1}^{6} T_j \phi_j(x, y).$$ \hfill (8.239)

Define the residual

$$\frac{\partial^2 \hat{T}}{\partial x^2} + \frac{\partial^2 \hat{T}}{\partial y^2} = R(x, y).$$ \hfill (8.240)

From the Galerkin method we have

$$\int_\Omega R(x, y) \phi_i(x, y) \, d\Omega = 0 \quad i = 1, \ldots, 6.$$ \hfill (8.241)

Expanding $R(x, y)$ we have

$$\int_\Omega \left( \frac{\partial^2 \hat{T}}{\partial x^2} + \frac{\partial^2 \hat{T}}{\partial y^2} \right) \phi_i(x, y) \, d\Omega = 0 \quad i = 1, \ldots, 6.$$ \hfill (8.242)

Substitution of 8.239 into 8.242 gives

$$\sum_{j=1}^{6} T_j \int_\Omega \left( \frac{\partial^2 \phi_j(x, y)}{\partial x^2} \phi_i(\xi, \eta) + \frac{\partial^2 \phi_j(x, y)}{\partial y^2} \phi_i(\xi, \eta) \right) \, d\Omega = 0 \quad i = 1, \ldots, 6 \quad (8.243)$$

We can now apply Green’s theorem to modify the second derivative terms and obtain

$$- \sum_{j=1}^{6} T_j \int_\Omega \left( \frac{\partial \phi_j(x, y)}{\partial x} \frac{\partial \phi_i(x, y)}{\partial x} + \frac{\partial \phi_j(x, y)}{\partial y} \frac{\partial \phi_i(x, y)}{\partial y} \right) \, d\Omega + \int_{\partial \Omega} \frac{\partial \hat{T}}{\partial n} \phi_i(x, y) \, dl = 0 \quad i = 1, \ldots, 6.$$ \hfill (8.244)

This set of equations can be written in global coordinates as
Since we know from the boundary conditions that
\[ T_1 = T_2 = 6 \quad (8.246) \]
and
\[ T_5 = T_6 = 12 \quad (8.247) \]
we can reduce the coefficient matrix to give
\[
\begin{bmatrix}
a_{33} & a_{34} \\
a_{43} & a_{44}
\end{bmatrix}
\begin{bmatrix}
T_3 \\
T_4
\end{bmatrix}
= \begin{bmatrix}
f_3 \\
f_4
\end{bmatrix}
- \begin{bmatrix}
a_{31}T_1 + a_{32}T_2 + a_{35}T_5 + a_{36}T_6 \\
a_{41}T_1 + a_{42}T_2 + a_{45}T_5 + a_{46}T_6
\end{bmatrix}. \quad (8.248)
\]
Although it may not be obvious, we can show from symmetry that
\[ a_{34} = a_{43} \quad (8.249) \]
and from geometry that
\[
\begin{align*}
a_{33} &= a_{44} \quad (8.250) \\
a_{31} &= a_{42} \quad (8.251) \\
a_{35} &= a_{46} \quad (8.252) \\
a_{36} &= a_{45} \quad (8.253) \\
a_{32} &= a_{41} \quad (8.254) \\
a_{35} &= a_{31} \quad (8.255) \\
a_{42} &= a_{46} \quad (8.256) \\
a_{36} &= a_{32}. \quad (8.257)
\end{align*}
\]
With these considerations in mind, we are left with the coefficients

\[ a_{31}, a_{32}, a_{33}, a_{34} \] \hspace{1cm} (8.258)

Consider coefficient \( a_{31} \):

\[ a_{31} = \int_{\Omega} \left( \frac{\partial \phi_1}{\partial x} \frac{\partial \phi_3}{\partial x} + \frac{\partial \phi_1}{\partial y} \frac{\partial \phi_3}{\partial y} \right) d\Omega. \] \hspace{1cm} (8.259)

The integration for this coefficient is restricted to element 1 because nodes 1 and 3 are only connected in element 1 as seen in Fig. 8.19. We now do the integration over only element 1 and we do this in the local coordinate system. The process is as follows.

**Transform Integrals to Natural Coordinate System**

The relationship between the natural and global coordinate systems for the first element are provided in the following table:

<table>
<thead>
<tr>
<th>Local</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 or ((-1, -1))</td>
<td>1</td>
</tr>
<tr>
<td>2 or ((1, -1))</td>
<td>3</td>
</tr>
<tr>
<td>3 or ((1, 1))</td>
<td>4</td>
</tr>
</tbody>
</table>
| 4 or \((-1, 1)\) | 2 | \hspace{1cm} (8.260)

Consider, for element 1 in natural coordinates, the integral (see Section 8.3.3 on page 219 above for insight into how this is done)

\[ \int_{\Omega_{e=1}} \left( \frac{\partial \phi_1 (\xi, \eta)}{\partial x} \frac{\partial \phi_2 (\xi, \eta)}{\partial x} + \frac{\partial \phi_1 (\xi, \eta)}{\partial y} \frac{\partial \phi_2 (\xi, \eta)}{\partial y} \right) d\Omega_e. \] \hspace{1cm} (8.261)

We need functions of the form \( \frac{\partial \phi_i (\xi, \eta)}{\partial \xi} \) defined in natural coordinates. We start with the relationship derived from the chain rule, that is,

\[ \left[ \begin{array}{c} \frac{\partial \phi_1 (\xi, \eta)}{\partial \xi} \\ \frac{\partial \phi_1 (\xi, \eta)}{\partial \eta} \end{array} \right] = \left[ \begin{array}{cc} \frac{\partial x}{\partial \xi} & \frac{\partial x}{\partial \eta} \\ \frac{\partial y}{\partial \xi} & \frac{\partial y}{\partial \eta} \end{array} \right] \left[ \begin{array}{c} \frac{\partial \phi_1 (\xi, \eta)}{\partial \xi} \\ \frac{\partial \phi_1 (\xi, \eta)}{\partial \eta} \end{array} \right]. \] \hspace{1cm} (8.262)

From Eq. (8.229) on page 222, reproduced below, we can write the coefficient matrix in Eq. (8.262) as

\[ \left[ \begin{array}{cc} \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{array} \right] = \frac{1}{4} \left[ \begin{array}{cccc} -1 + \eta & 1 - \eta & 1 + \eta & -1 - \eta \\ -1 + \xi & -1 - \xi & 1 + \xi & 1 - \xi \end{array} \right] \left[ \begin{array}{cccc} x_1 & y_1 & x_2 & y_2 \\ x_3 & y_3 & x_4 & y_4 \end{array} \right]. \] \hspace{1cm} (8.263)

So, for the particular case of \( \phi_1 (\xi, \eta) \) we substitute Eq. (8.263) into Eq. (8.262) to obtain

\[ \left[ \begin{array}{c} \frac{\partial \phi_1 (\xi, \eta)}{\partial \xi} \\ \frac{\partial \phi_1 (\xi, \eta)}{\partial \eta} \end{array} \right] = \frac{1}{4} \left[ \begin{array}{cccc} -1 + \eta & 1 - \eta & 1 + \eta & -1 - \eta \\ -1 + \xi & -1 - \xi & 1 + \xi & 1 - \xi \end{array} \right] \left[ \begin{array}{cccc} 0 & 0 & 10 & 0 \\ 10 & 5 & 0 & 5 \end{array} \right] \frac{\partial \phi_1 (\xi, \eta)}{\partial \eta}. \] \hspace{1cm} (8.264)
or

\[
\begin{bmatrix}
\frac{\partial \phi_1(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_1(\xi, \eta)}{\partial \eta}
\end{bmatrix} = \frac{1}{4} \begin{bmatrix}
(1 - \eta)(10) + (1 + \eta)(10) & (1 + \eta)(5) + (-1 - \eta)(5) \\
(-1 - \xi)(10) + (1 + \xi)(10) & (1 + \xi)(5) + (1 - \xi)(5)
\end{bmatrix} \begin{bmatrix}
\frac{\partial \phi_1(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_1(\xi, \eta)}{\partial \eta}
\end{bmatrix} \quad (8.265)
\]

Similarly we have

\[
\begin{bmatrix}
\frac{\partial \phi_2(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_2(\xi, \eta)}{\partial \eta}
\end{bmatrix} = \frac{1}{4} \begin{bmatrix}
20 & 0 \\
0 & 10
\end{bmatrix} \begin{bmatrix}
\frac{\partial \phi_2(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_2(\xi, \eta)}{\partial \eta}
\end{bmatrix} . \quad (8.266)
\]

Taking the inverse of the Jacobian matrix we have

\[
\begin{bmatrix}
\frac{\partial \phi_1(\xi, \eta)}{\partial x} \\
\frac{\partial \phi_1(\xi, \eta)}{\partial y}
\end{bmatrix} = \frac{16}{200} \begin{bmatrix}
\frac{10}{4} & 0 \\
0 & \frac{20}{4}
\end{bmatrix} \begin{bmatrix}
\frac{\partial \phi_1(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_1(\xi, \eta)}{\partial \eta}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\frac{1}{5} & 0 \\
0 & \frac{2}{5}
\end{bmatrix} \begin{bmatrix}
\frac{\partial \phi_1(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_1(\xi, \eta)}{\partial \eta}
\end{bmatrix} \quad (8.268)
\]

and

\[
\begin{bmatrix}
\frac{\partial \phi_2(\xi, \eta)}{\partial x} \\
\frac{\partial \phi_2(\xi, \eta)}{\partial y}
\end{bmatrix} = \begin{bmatrix}
\frac{1}{5} & 0 \\
0 & \frac{2}{5}
\end{bmatrix} \begin{bmatrix}
\frac{\partial \phi_2(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_2(\xi, \eta)}{\partial \eta}
\end{bmatrix} . \quad (8.269)
\]

This is the relationship we need to transform our derivatives in Eq. (8.261) so they are in the form we need.

To see how this is done, we return to our integral of interest, that is

\[
\int_{\Omega_{e=1}} \left( \frac{\partial \phi_1(\xi, \eta)}{\partial x} \frac{\partial \phi_2(\xi, \eta)}{\partial x} + \frac{\partial \phi_1(\xi, \eta)}{\partial y} \frac{\partial \phi_2(\xi, \eta)}{\partial y} \right) \, d\Omega_e
\]

which becomes, using Eqs. (8.268) and (8.269) as follows:

\[
\int_{\Omega_{e=1}} \left( \frac{\partial \phi_1(\xi, \eta)}{\partial x} \frac{\partial \phi_2(\xi, \eta)}{\partial x} + \frac{\partial \phi_1(\xi, \eta)}{\partial y} \frac{\partial \phi_2(\xi, \eta)}{\partial y} \right) \, d\Omega_e
\]

\[
= \int_{-1}^{1} \int_{-1}^{1} \left[ \left( \frac{1}{5} \frac{\partial \phi_1(\xi, \eta)}{\partial \xi} \right) \left( \frac{1}{5} \frac{\partial \phi_2(\xi, \eta)}{\partial \xi} \right) \\
+ \left( \frac{2}{5} \frac{\partial \phi_1(\xi, \eta)}{\partial \eta} \right) \left( \frac{2}{5} \frac{\partial \phi_2(\xi, \eta)}{\partial \eta} \right) \right] \det[J] \, d\xi d\eta \quad (8.271)
\]

and \( \det[J] \) is given by
\[
\det \frac{1}{4} \begin{bmatrix} 20 & 0 \\ 0 & 10 \end{bmatrix} = \det \begin{bmatrix} 5 & 0 \\ 0 & \frac{5}{2} \end{bmatrix} = \frac{25}{2}.
\] (8.272)

**Perform Integrations**

We now use Gauss quadrature (see Section 3.1.6 on page 59) to evaluate integrals of the form found in Eq. (8.271). The Gauss points are at

\[
b_i (\xi) = \pm \frac{1}{\sqrt{3}}, \quad b_i (\eta) = \pm \frac{1}{\sqrt{3}}
\] (8.273)

where \(\frac{1}{\sqrt{3}} = 0.5773\), and the weighting function is \(H_i = 1.0\). Consider the first term in the integral on the right-hand side of Eq. (8.271). From the definitions of \(\phi_1 (\xi, \eta)\) and \(\phi_2 (\xi, \eta)\) given by

\[
\phi_1 = \left[ \left( \frac{1 - \xi}{2} \right) \left( \frac{1 - \eta}{2} \right) \right]
\] (8.274)

and

\[
\phi_2 = \left[ \left( \frac{1 + \xi}{2} \right) \left( \frac{1 - \eta}{2} \right) \right]
\] (8.275)

we have for the first term in Eq. (8.271),

\[
\begin{align*}
\int_{-1}^{1} \int_{-1}^{1} \left( \frac{1}{5} \frac{\partial \phi_1 (\xi, \eta)}{\partial \xi} \right) \left( \frac{1}{5} \frac{\partial \phi_2 (\xi, \eta)}{\partial \xi} \right) \det [J] \, d\xi \, d\eta \\
= \sum_{\xi = \pm \frac{1}{\sqrt{3}}} \sum_{\eta = \pm \frac{1}{\sqrt{3}}} \left[ \left( \frac{1}{5} \left[ - \left( \frac{1 - \eta}{4} \right) \right] \right) \left( \frac{1}{5} \left[ \left( \frac{1 - \eta}{4} \right) \right] \right) \right] \frac{25}{2}
\end{align*}
\]

\[
= 2 \sum_{\xi = \pm \frac{1}{\sqrt{3}}} \left( \frac{1}{25} \right) \left( \frac{1}{16} \right) (1 - 2\eta + \eta^2) \left( \frac{25}{2} \right)
\]

\[
= 2 \left\{ - \left( \frac{1}{25} \right) \left( \frac{1}{16} \right) \left[ 1 - 2\left( \frac{1}{\sqrt{3}} \right)^2 + \left( \frac{1}{\sqrt{3}} \right)^2 \right] - \left( \frac{1}{25} \right) \left( \frac{1}{16} \right) \left[ 1 - 2\left( \frac{1}{\sqrt{3}} \right)^2 + \left( \frac{1}{\sqrt{3}} \right)^2 \right] \right\} \left( \frac{25}{2} \right)
\]

\[
= -2 \left( \frac{1}{25} \right) \left( \frac{1}{16} \right) \left[ 1 + \frac{2}{\sqrt{3}} + \frac{1}{3} \right] \left[ 1 + \frac{2}{\sqrt{3}} + \frac{1}{3} \right] \left( \frac{25}{2} \right)
\]

\[
= -2 \left( \frac{1}{25} \right) \left( \frac{1}{16} \right) \left( 2 + \frac{2}{3} \right) \left( \frac{25}{2} \right)
\]

\[
= -\frac{2}{16} \left( \frac{8}{3} \right) \left( \frac{1}{2} \right)
\]

\[
= -\frac{1}{6}.
\] (8.276)
Note that the notation \( \sum_{\xi=\pm \frac{1}{\sqrt{3}}} \) means that you substitute in the values \( \xi = \pm \frac{1}{\sqrt{3}} \) into the argument sequentially and sum the result. Thus there will be four terms in the summation on the right-hand side of the above equation. Now consider the second term on the right-hand side of Eq. (8.271).

\[
\int_{-1}^{1} \int_{-1}^{1} \left[ \left( \frac{2}{5} \frac{\partial \phi_1 (\xi, \eta)}{\partial \eta} + \frac{2}{5} \frac{\partial \phi_2 (\xi, \eta)}{\partial \eta} \right) \right] \det [J] d\xi d\eta
= \sum_{\eta=\pm \frac{1}{\sqrt{3}}} \sum_{\xi=\pm \frac{1}{\sqrt{3}}} \left[ \left( \frac{2}{5} \left[ -\left( \frac{1-\xi}{4} \right) \right] \right) \left( \frac{2}{5} \left[ -\left( \frac{1+\xi}{4} \right) \right] \right) \frac{25}{2} \right]
= 2 \sum_{\eta=\pm \frac{1}{\sqrt{3}}} \left( \frac{4}{25} \right) \left( \frac{1}{16} \right) (1-\xi^2) \left( \frac{25}{2} \right)
= 2 \left( \frac{4}{25} \right) \left( \frac{1}{16} \right) \left[ \left( 1 - \frac{1}{3} \right) + \left( 1 - \frac{1}{3} \right) \right] \left( \frac{25}{2} \right)
= 2 \left( \frac{4}{25} \right) \left( \frac{1}{3} \right) \left( \frac{1}{2} \right)
= \frac{2}{6}.
\] (8.277)

The total term is the sum of \(-\frac{1}{6} + \frac{2}{6} = \frac{1}{6}\). We now have the value for the coefficient \((13)\) in the local coefficient matrix for element 1.

**Matrix Assembly**

If we evaluate all the integral terms for both elements, we can obtain the element coefficient matrices.

To obtain the global coefficient matrix we need a transformation table for the global and local nodal numbers. In the top row we have the global node numbers. In the second row, the local element node number of each node value for element 1 is provided. For example, the local node number in element 1 for global node 2 is 4. The same information for element 2 is found in the last row.

<table>
<thead>
<tr>
<th>Global</th>
<th>Local</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>Element</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>

Elements 1 and 2 have the same coefficient values since they have the same geometry and properties, that is,

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} \\
  a_{21} & a_{22} & a_{23} & a_{24} \\
  a_{31} & a_{32} & a_{33} & a_{34} \\
  a_{41} & a_{42} & a_{43} & a_{44}
\end{bmatrix}
= \begin{bmatrix}
  5/6 & 1/6 & -5/12 & -7/12 \\
  1/6 & 5/6 & -7/12 & -5/12 \\
  -5/12 & -7/12 & 5/6 & 1/6 \\
  -7/12 & -5/12 & 1/6 & 5/6
\end{bmatrix}.
\] (8.279)

The global matrix is obtained by combining information from the two element matrices and the transformation table. As an example, consider the global coefficient \((b_{11})\). This coefficient is according to the table, the \((a_{11})\) coefficient of the element.
coefficient matrix for element 1; this value is 5/6 and this value appears as the \((b_{11})\) coefficient in the global matrix given in Eq. (8.280). Now let us consider a more interesting case. Node 3 in Fig. 8.19 lies at the interface of the two elements. Thus this coefficient in the global matrix draws information from both of these elements. We see this in the table because global node 3 is the same as local node 2 in element 1 it is the same as local node 1 in element 2. So, to build the global coefficient \((b_{33})\) we need to add the \((a_{22})\) coefficient from element 1 to the \((a_{11})\) coefficient from element 2. We obtain 5/6 + 5/6 as is seen in the \((b_{33})\) location of the matrix in Eq. (8.280).

As a final example we take the most complicated situation, the formation of global element coefficient \((b_{34})\). This coefficient involves information from two nodes, each located at the interface of two elements. From the table we see that in the first element, global node \((b_{34})\) is equivalent to element node \((a_{23})\) in element 1 and \((a_{14})\) in element 2. Thus we need to add these two values together to obtain the global coefficient \((b_{34})\) and so we obtain \(-7/12 -7/12\) as seen in coefficient location \((b_{34})\).

Note that for this second-order equation both the global \([B]\) and local \([A]\) coefficient matrices are symmetric and the rows and columns sum to zero.

\[
\begin{pmatrix}
  b_{11} & b_{12} & b_{13} & b_{14} & b_{15} & b_{16} \\
  b_{21} & b_{22} & b_{23} & b_{24} & b_{25} & b_{26} \\
  b_{31} & b_{32} & b_{33} & b_{34} & b_{35} & b_{36} \\
  b_{41} & b_{42} & b_{43} & b_{44} & b_{45} & b_{46} \\
  b_{51} & b_{52} & b_{53} & b_{54} & b_{55} & b_{56} \\
  b_{61} & b_{62} & b_{63} & b_{64} & b_{65} & b_{66}
\end{pmatrix} = \begin{pmatrix}
  5/6 & -7/12 & 1/6 & -5/12 & 0 & 0 \\
  -7/12 & 5/6 & -5/12 & 1/6 & 0 & 0 \\
  1/6 & -5/12 & 5/6 + 5/6 & -7/12 & -7/12 & 1/6 \\
  -5/12 & 1/6 & -7/12 & 5/6 & -7/12 & -7/12 \\
  0 & 0 & 1/6 & -7/12 & 5/6 & -7/12 \\
  0 & 0 & -5/12 & 5/6 & -7/12 & 5/6 \\
\end{pmatrix}
\]

\(8.280\)

The complete matrix equation is

\[
\begin{pmatrix}
  5/6 & -7/12 & 1/6 & -5/12 & 0 & 0 \\
  -7/12 & 5/6 & -5/12 & 1/6 & 0 & 0 \\
  1/6 & -5/12 & 5/6 + 5/6 & -7/12 & -7/12 & 1/6 \\
  -5/12 & 1/6 & -7/12 & 5/6 & -7/12 & -7/12 \\
  0 & 0 & 1/6 & -7/12 & 5/6 & -7/12 \\
  0 & 0 & -5/12 & 5/6 & -7/12 & 5/6 \\
\end{pmatrix} \begin{pmatrix}
T_1 \\
T_2 \\
T_3 \\
T_4 \\
T_5 \\
T_6
\end{pmatrix} = \begin{pmatrix}
f_1 \\
f_2 \\
f_3 \\
f_4 \\
f_5 \\
f_6
\end{pmatrix}
\]

\(8.281\)

where

\[f_i = -\int_{\partial\Omega} \frac{\partial T}{\partial n} \phi_i(x, y) \, dl.\]

\(8.282\)

Introducing the boundary conditions we have

\[
\begin{pmatrix}
  5/7 & -7/12 & 1/6 & -5/12 & 0 & 0 \\
  -7/12 & 5/6 & -5/12 & 1/6 & 0 & 0 \\
  1/6 & -5/12 & 5/6 + 5/6 & -7/12 & -7/12 & 1/6 \\
  -5/12 & 1/6 & -7/12 & 5/6 & -7/12 & -7/12 \\
  0 & 0 & 1/6 & -7/12 & 5/6 & -7/12 \\
  0 & 0 & -5/12 & 5/6 & -7/12 & 5/6 \\
\end{pmatrix} \begin{pmatrix}
6 \\
T_3 \\
T_4 \\
12 \\
T_5 \\
12
\end{pmatrix} = \begin{pmatrix}
f_1 \\
f_2 \\
f_3 \\
f_4 \\
f_5 \\
f_6
\end{pmatrix}
\]

\(8.283\)
which after bringing information to the right-hand side of the equation simplifies to

\[
- \begin{bmatrix}
\frac{5}{6} + \frac{5}{6} & -7/12 - 7/12 \\
-7/12 - 7/12 & \frac{5}{6} + \frac{5}{6}
\end{bmatrix}
\begin{bmatrix}
T_3 \\
T_4
\end{bmatrix}
= \begin{bmatrix}
(1/6)(6) + (-5/12)(6) + (1/6)(12) + (-5/12)(12) \\
(-5/12)(6) + (1/6)(6) + (-5/12)(12) + (1/6)(12)
\end{bmatrix}
\]  

\(8.284\)

\[- \begin{bmatrix}
\frac{5}{3} & -7/6 \\
-7/6 & \frac{5}{3}
\end{bmatrix}
\begin{bmatrix}
T_3 \\
T_4
\end{bmatrix}
= \begin{bmatrix}
9/2 \\
9/2
\end{bmatrix}.
\]  

\(8.285\)

Solving for the unknown vector we get (on inversion of the coefficient matrix)

\[
\begin{bmatrix}
T_3 \\
T_4
\end{bmatrix}
= \frac{1}{(5/3)(5/3) - (7/6)(7/6)} \begin{bmatrix}
5/3 & 7/6 \\
7/6 & 5/3
\end{bmatrix}
\begin{bmatrix}
9/2 \\
9/2
\end{bmatrix}
\]

\[
= \frac{36}{51} \begin{bmatrix}
(5/3 + 7/6)(9/2) \\
(7/6 + 5/3)(9/2)
\end{bmatrix}
\]

\[
= \frac{3}{51} \begin{bmatrix}
(17/1)(9/1) \\
(17/1)(9/1)
\end{bmatrix}
\]

\[
= \begin{bmatrix}
9 \\
9
\end{bmatrix}
\]  

\(8.286\)

which is the exact solution.

### 8.4 Chapter Summary

In this chapter we introduce the Galerkin finite element method. We begin with rectangular elements. The approach used for rectangles is next extended to triangular finite elements, a method that is somewhat more abstract than the case of rectangular elements. The rectangular finite element formulation was next generalized to the case of isoparametric finite element that allowed for elements that are squares in a local coordinate system but can be deformed rectangles or trapezoids in the global coordinate system. The integrations inherent in the finite element method are easily performed in this local coordinate system.

### 8.5 Problems

1. Given a triangle with nodes at \((0,0)\), \((2,0)\) and \((1,2)\) as shown below in Fig. 8.20 and the values \(u(0,0) = 1\), \(u(2,0) = 2\), and \(u(1,2) = 0\), determine the value of \(u(1,1)\) and show that it is 1.0 This can be determined from the following two relationships:

\[
\hat{u}(x,y) = \sum_{i=1}^{3} u_i L_i(x,y)  
\]  

\(8.287\)

\[
\sum_{i=1}^{3} L_i(x,y) = 1  
\]  

\(8.288\)

where \(L_i(x,y)\) is a triangular basis function.
2. Given the following equation that is used in the isoparametric formulation,

\[
\begin{bmatrix}
\frac{\partial \phi_j(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_j(\xi, \eta)}{\partial \eta}
\end{bmatrix}
= \begin{bmatrix}
\frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\
\frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta}
\end{bmatrix}
\begin{bmatrix}
\frac{\partial \phi_j(\xi, \eta)}{\partial \xi} \\
\frac{\partial \phi_j(\xi, \eta)}{\partial \eta}
\end{bmatrix}
\]

(8.289)

and the information in Figs. 8.21 and 8.22, calculate the value of \( \frac{\partial x}{\partial \xi}, \frac{\partial y}{\partial \xi}, \frac{\partial x}{\partial \eta}, \) and \( \frac{\partial y}{\partial \eta} \), and show that they are as follows:

\[
\begin{bmatrix}
\frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\
\frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta}
\end{bmatrix} = \begin{bmatrix}
5 & 0 \\
0 & 2.5
\end{bmatrix}
\]

(8.290)

Hint: You can do this problem just through careful observation of the figures below and the properties of basis functions, but just in case you need them here are the equations for the basis functions: \( \phi_j(x, y) = \phi_j(x) \phi_j(y) \) where \( \phi_j(x) = \frac{x_j+1-x}{\Delta x} \) and \( \phi_j(\xi, \eta) = \phi_j(\xi) \phi_j(\eta) \) where \( \phi_j(\xi) = \frac{\xi_j+1-\xi}{2} \).

3. Consider the finite element array found in Fig. 8.5 the following figure. Assume the global matrix for this problem is

\[
\begin{bmatrix}
1 & 1 & 0 & 2 \\
1 & 3 & 2 & 5 \\
0 & 2 & 2 & 4 \\
2 & 5 & 4 & 4
\end{bmatrix}
\]

and the element matrix for element \( e = 1 \) is

\[
\begin{bmatrix}
1 & 1 & 2 \\
1 & 2 & 2 \\
2 & 2 & 1
\end{bmatrix}
\]

(8.291)

The transformation table is
Figure 8.21: Explanation sketch for problem 2.

Figure 8.22: Explanation sketch for problem 2.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Element one</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Element two</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please fill in the values for element matrix $e = 2$

\[
\begin{bmatrix}
1 & 2 & 3 & 4 \\
1 & 2 & 3 & \\
1 & 2 & 3 & \\
\end{bmatrix}
\]

(8.292)
4. Given a triangle as illustrated in Fig. 8.5, with nodes at \((0,0), (2,0),\) and \((1,2)\) as shown below and the values \(u(0,0) = 1, u(2,0) = 2,\) and \(u(1,2) = 0,\) and the basis functions

\[
L_I(x,y) = \frac{x(y_{I+1} - y_{I+2}) + y(x_{I+2} - x_{I+1}) + (x_{I+1}y_{I+2} - x_{I+2}y_{I+1})}{\det [P]} \tag{8.293}
\]

where

\[
A = \frac{1}{2} \det [P] \tag{8.294}
\]

and \(A = 2\) is the area of the triangle, show, using the above basis functions, that the value of \(\frac{\partial u(x,y)}{\partial x}\) at the location \((x,y) = (0.5,0.5)\) is 0.5.

5. Given the finite element mesh in Fig. 8.5, locate and place the global nodal numbers on the figure (in the circles). In the following table, the global numbers are in bold and the element numbers are in italics. Remember that local element numbering (the \(I, I+1, I+2\)) is counter-clockwise.
7. Consider the equation

\[ \frac{\partial}{\partial x}u + \frac{\partial}{\partial y}u = 0 \quad (8.295) \]

defined in the one-element domain shown in Fig. 8.23 with boundary conditions

\[ u(0,0) = 1 \quad (8.296) \]
\[ u(1,0) = 0. \]

Solve for the value at node 3, namely \( u(0,1) \), or equivalently \( u_3 \), using a Galerkin finite element method.

Steps

1. Write the approximation for \( u(x,y) \), that is, \( \hat{u}(x,y) = \sum_{j=1}^{3} u_j \phi_j(x,y) \).
2. Generate the weighted residual equations (there are three).
3. Substitute for \( \hat{u}(x,y) \) in the weighted residual equation.
4. Write the matrix equation symbolically, that is the global matrix \((3 \times 3)\) in terms of the integral such as \( \int_{\Omega} \frac{\partial}{\partial x} \phi \, da \), the unknown vector of values of \( u_j \) and the right-hand side (which at this stage is zero).
5. Impose the two boundary conditions (you will now have one unknown, \( u_3 \)) and note that you will now have integrals in the \((1 \times 1)\) global matrix and on the right hand side (two of them).
6. Evaluate the integrals using the geometry shown in the figure to define the derivatives of \( \phi_j \) and the relationship \( \int_{\Omega} \phi_i \, da = \frac{A}{3} \) (you actually don’t need this last relationship, but in case you don’t see why, here it is).
7. You now have one equation in one unknown, \( u_3 \). Solve for it.

8 Check your answer by determining whether the derivatives of \( u(x, y) \) satisfy the original differential equation.

Figure 8.23: One-element domain with the basis function at node one illustrated.
Bibliography


Chapter 9

Finite Volume Approximation in Two Space

9.1 Finite Volume Formulation

In this section we will extend the finite volume formulation we introduced in Section 5.1 on page 84 to two-dimensional problems. The system we will consider is one that allows for a cross derivative such as $\frac{\partial^2 u(x,y)}{\partial x \partial y}$, namely

$$
a_{xx} \frac{\partial^2 u(x,y)}{\partial x^2} + a_{yy} \frac{\partial^2 u(x,y)}{\partial y^2} + a_{xy} \frac{\partial^2 u(x,y)}{\partial x \partial y} + a_{yx} \frac{\partial^2 u(x,y)}{\partial y \partial x} + u(x,y) = \rho(x,y) \quad x,y \in \Omega$$

where $\rho(x,y)$ is the source term.

$$
+ \kappa u(x,y) = 0 \quad x,y \in \Omega
$$

$$
u(x,y) = \nu_0(x,y) \quad x,y \in \partial \Omega_D
$$

$$
\frac{\partial u(x,y)}{\partial n} = \nu_1(x,y) \quad x,y \in \partial \Omega_N
$$

$$
\partial \Omega_D + \partial \Omega_N = \partial \Omega
$$

where, as earlier, $\partial \Omega_D$ and $\partial \Omega_N$ are the segments of the boundary $\partial \Omega$ of the domain $\Omega$ along which Dirichlet and Neumann boundary conditions are defined respectively.

The formulation of the finite volume method in two space dimensions can be done using more than one approach. The one we have selected is chosen because it is consistent with the method of weighted residuals which we have used as the intellectual glue to hold together the various methods we have introduced thus far. With this in mind, as in the case of earlier methods, we begin with a representation of $u(x,y)$ as

$$
u(x,y) \approx \hat{u}(x,y) = \sum_{j=1}^{N} u_j \phi_j(x,y)
$$

where, as in Section 8.2 on page 195, we use basis functions defined on triangular elements.

A typical finite element arrangement is seen in Fig. 9.1. The black dots are the nodes. The small circles are the circumcenters of the triangles. The circumcenter is the location of the intersection of the perpendicular bisectors of each side. It is also the center of a circle that passes through all three of the nodes of the triangular element. We will make reference to circumcenters in the next paragraph.
To begin the development of the finite volume approximation, a series of polygons is constructed within the array of elements as seen in Fig. 9.2. The polygons are created by connecting the circumcenters of each triangle. Note that the sides of the polygons are orthogonal to the sides of the triangles and are continuous across them. The polygons, in the absence of the triangles, are shown in Fig. 9.3. We will now discuss the rationale for the various steps just presented and we will explain, via two examples, how to employ the resulting finite volume formulation.

![Diagram](image)

Figure 9.1: Domain of interest is subdivided into elements as part of the finite volume formulation presentation. The small circles are the approximate locations of the circumcenters of the triangles.

Once we have the finite element basis functions defined, as presented in Eq. (9.6) and discussed in more detail later, the next step is to write the weighted residual formulation, that is

$$\int_{\Omega} R(x, y) w_i(x, y) \, d\Omega = 0 \quad i = 1, 2, \ldots, N$$  \hspace{1cm} (9.7)

which contains the weighting function \(w_i(x, y)\). The choice of the weighting function for the two-dimensional finite volume method is a direct extension of the one-dimensional case discussed in Section 5.4 on page 112.

To visualize our choice, imagine extending a line of unit length vertically at the circumcenter of each triangle as denoted by the letter C in Fig. 9.5. Connect the tops of the resulting lines to form the polygon at a height of 1 unit (this is the surface B in Fig. 9.5). Finally stretch a hypothetical membrane over the polygon on the top and then enclose each of the vertical lines just constructed with the membrane. The resulting ‘tower’ is the needed weighting function \(w_i(x, y)\) and is consistent with the finite element finite volume array shown in Fig. 9.6. The projection of the weighting function onto the \(x-y\) plane is illustrated in Figs. 9.5 and 9.6. We will learn why we have selected this functional form for the weighting function a little later.

As we have done in earlier examples, we now expand the weighted residual formu-
9.1. FINITE VOLUME FORMULATION

Figure 9.2: Polygonal finite volumes generated from the finite element array shown in Fig. 9.1.

Now we apply Green’s theorem to the second-order terms in Eq. (9.8) and we get

$$
\begin{align*}
\int_{\Omega} \left( a_{xx} \frac{\partial^2 \hat{u}(x,y)}{\partial x^2} + a_{yy} \frac{\partial^2 \hat{u}(x,y)}{\partial y^2} \right) d\Omega + \\
\int_{\Omega} \left( a_{xy} \frac{\partial^2 \hat{u}(x,y)}{\partial x \partial y} + a_{yx} \frac{\partial^2 \hat{u}(x,y)}{\partial y \partial x} + \kappa \hat{u}(x,y) - Qw_i(x,y) \right) d\Omega = 0 \quad i = 1, 2, ..., N.
\end{align*}
$$

(9.8)

where $n_x$ and $n_y$ are the components of the vector $n$ that is normal to $\Omega$ in the $x$ and $y$ directions respectively.

Consider now the information provided in Fig. 9.4. Here you see a typical finite element with the coordinates $x, y, n,$ and $\tau$ defined. Note that $n$ is orthogonal to the side of the finite volume and $\tau$ is tangential to it.

Next we chain out the $w_i$ derivatives with respect to $\tau$ and $n$. To see how this is done, recall that by definition
The unit vectors in the $x$ and $y$ coordinate directions are given by $i$ and $j$ respectively. We can also use the definition of the gradient operator to write it in the normal-tangential coordinate system

$$\frac{\partial}{\partial n} w_i (x,y) = \nabla w_i (x,y) \cdot n$$

and

$$\frac{\partial}{\partial \tau} w_i (x,y) = \nabla w_i (x,y) \cdot \tau$$

where, in this case

$$\nabla w_i (x,y) = \frac{\partial}{\partial n} w_i (x,y) n + \frac{\partial}{\partial \tau} w_i (x,y) \tau$$

and $n$ and $\tau$ are the unit vectors in the $n$, and $\tau$ coordinate directions respectively.

Now we expand this relationship to give

$$\frac{\partial}{\partial x} w_i (x,y) = \left( \frac{\partial}{\partial n} w_i (x,y) n + \frac{\partial}{\partial \tau} w_i (x,y) \tau \right) \cdot i.$$  \hspace{1cm} (9.15)

We can now rewrite this expression as

$$\frac{\partial}{\partial x} w_i (x,y) = \left( \frac{\partial}{\partial n} w_i (x,y) n \cdot i + \frac{\partial}{\partial \tau} w_i (x,y) \tau \cdot i \right).$$  \hspace{1cm} (9.16)

Similarly, for $\frac{\partial}{\partial y} w_i (x,y)$ we obtain

$$\frac{\partial}{\partial y} w_i (x,y) = \left( \frac{\partial}{\partial n} w_i (x,y) n \cdot j + \frac{\partial}{\partial \tau} w_i (x,y) \tau \cdot j \right).$$  \hspace{1cm} (9.17)
Figure 9.4: Coordinate system relationship associated with finite volume formulation. Note that the positive angle for $\theta$ is counter-clockwise.

As noted earlier and illustrated in Figs. 9.5 and 9.6, the weighting function $w_i(x, y)$ is a constant in the $\tau$ direction such that the terms of the form $\frac{\partial}{\partial \tau} w_i(x, y)$ are zero. Using this property, the above equations become

$$\frac{\partial}{\partial x} w_i(x, y) = \frac{\partial}{\partial \tau} w_i(x, y) \mathbf{n} \cdot \mathbf{i} = -\delta(x, y) |_{\partial \Omega}, \mathbf{n} \cdot \mathbf{i}$$

(9.18)

and

$$\frac{\partial}{\partial y} w_i(x, y) = \frac{\partial}{\partial \tau} w_i(x, y) \mathbf{n} \cdot \mathbf{j} = -\delta(x, y) |_{\partial \Omega}, \mathbf{n} \cdot \mathbf{j}$$

(9.19)

where $\delta(x, y) |_{\partial \Omega}$ is the Dirac delta function defined along $\partial \Omega$. 

Now we need to determine the product $\mathbf{n} \cdot \mathbf{j}$. Referring to Fig. 9.4 we see that the vector from $a$ to $c$ in the $n$ direction can be expressed as

$$\mathbf{n}_{ac} = (x_c - x_a, y_c - y_a).$$

(9.20)

To normalize this vector to obtain the unit vector $\mathbf{n}$, we divide by the length of $\mathbf{n}_{ac}$, that is

$$\mathbf{n} = \frac{(x_c - x_a, y_c - y_a)}{\sqrt{(x_c - x_a)^2 + (y_c - y_a)^2}}.$$

(9.21)

Thus we find the unit vector coordinates for $\mathbf{n}_x$ to be

$$\mathbf{n} \cdot \mathbf{i} \equiv n_x = \frac{x_c - x_a}{\sqrt{(x_c - x_a)^2 + (y_c - y_a)^2}}$$

(9.22)

and for $\mathbf{n}_y$ to be

$$\mathbf{n} \cdot \mathbf{j} \equiv n_y = \frac{y_c - y_a}{\sqrt{(x_c - x_a)^2 + (y_c - y_a)^2}}.$$

(9.23)

Note that the above calculations could be performed using the coordinates at the locations $c, m,$ and $n$ in lieu of $c, a,$ and $b$ with the same results.
Figure 9.5: The weighting function used in the finite volume formulation. The top of
the function is denoted by the dark-grey polygon.

We return to Eq. (9.9) and introduce the relationships developed above to obtain

\[
- \int_{\Omega} \left[ \left( \frac{a_{xx}}{\partial x} + \frac{a_{xy}}{\partial y} \right) (-\delta(x, y) \mid \partial \Omega_i) \right] \mathbf{n} \cdot \mathbf{i} \\
+ \int_{\Omega} \left[ \left( a_{yx} \frac{\partial \hat{u}(x, y)}{\partial x} + a_{yy} \frac{\partial \hat{u}(x, y)}{\partial y} \right) (-\delta(x, y) \mid \partial \Omega_i) \right] \mathbf{n} \cdot \mathbf{j} \\
+ \int_{\partial \Omega} \left[ \left( \frac{a_{xx}}{\partial x} + \frac{a_{xy}}{\partial y} \right) n_x \right. \\
+ \left( a_{yx} \frac{\partial \hat{u}(x, y)}{\partial x} + a_{yy} \frac{\partial \hat{u}(x, y)}{\partial y} \right) n_y \right] w_i(x, y) ds = 0 \\
i = 1, 2, \ldots, N
\]  

(9.24)

where \( \hat{u}(x, y) \) is presented in Eq. (9.6). From the definition of the Dirac delta function,
this equation can be rewritten as

\[
- \int_{\Omega} \left[ \left( \frac{a_{xx}}{\partial x} + \frac{a_{xy}}{\partial y} \right) (-\delta(x, y) \mid \partial \Omega_i) \right] \mathbf{n} \cdot \mathbf{i} \\
+ \int_{\Omega} \left[ \left( a_{yx} \frac{\partial \hat{u}(x, y)}{\partial x} + a_{yy} \frac{\partial \hat{u}(x, y)}{\partial y} \right) (-\delta(x, y) \mid \partial \Omega_i) \right] \mathbf{n} \cdot \mathbf{j} \\
+ \int_{\partial \Omega} \left[ \left( \frac{a_{xx}}{\partial x} + \frac{a_{xy}}{\partial y} \right) n_x \right. \\
+ \left( a_{yx} \frac{\partial \hat{u}(x, y)}{\partial x} + a_{yy} \frac{\partial \hat{u}(x, y)}{\partial y} \right) n_y \right] w_i(x, y) ds = 0 \\
i = 1, 2, \ldots, N
\]
9.1. **FINITE VOLUME FORMULATION**

Figure 9.6: Finite volume (gray) and finite elements with illustration of normal $n$ and tangential $\tau$ coordinate directions.

\[
\int_{\partial \Omega_i} \left[ \left( a_{xx} \frac{\partial \hat{u}(x,y)}{\partial x} + a_{xy} \frac{\partial \hat{u}(x,y)}{\partial y} \right) n_x + \left( a_{yx} \frac{\partial \hat{u}(x,y)}{\partial x} + a_{yy} \frac{\partial \hat{u}(x,y)}{\partial y} \right) n_y \right] ds \\
+ \int_\Omega (\kappa \hat{u}(x,y) - Q) w_i(x,y) d\Omega \\
+ \int_{\partial \Omega} \left[ \left( a_{xx} \frac{\partial \hat{u}(x,y)}{\partial x} + a_{xy} \frac{\partial \hat{u}(x,y)}{\partial y} \right) n_x \\
+ \left( a_{yx} \frac{\partial \hat{u}(x,y)}{\partial x} + a_{yy} \frac{\partial \hat{u}(x,y)}{\partial y} \right) n_y \right] w_i(x,y) ds = 0 \\
i = 1, 2, ..., N. \tag{9.25}
\]

Notice that integration is now over the finite volume perimeter and that there are two uses of the notation $n_x$ and $n_y$. One is associated with the finite volume $i$ and the other with the region $\Omega$.

As a final step, we introduce the triangular basis functions $\phi_j(x,y)$ to give the final form we will use in the following examples.
The above development may seem a bit abstract at this point, but how we use this information in application will be clearer after we work through the following two examples:

### 9.2 Finite Volume Example Problem 1

#### 9.2.1 Problem Definition

Consider the system of equations presented above modified to be descriptive of the problem presented in Fig. 9.7, namely

\[
\begin{align*}
2 \frac{\partial^2 u(x, y)}{\partial x^2} + \frac{\partial^2 u(x, y)}{\partial y^2} + \kappa u(x, y) - Q &= 0 \quad x, y \in \Omega, \\
u(x, y) &= 1 \quad x = 0, \quad y \in [0, 2], \\
u(x, y) &= 0 \quad x = 2, \quad y \in [0, 2], \\
\frac{\partial u(x, y)}{\partial n} &= 0 \quad x \in (0, 2), \quad y = 0, \\
\frac{\partial u(x, y)}{\partial n} &= 0 \quad x \in (0, 2), \quad y = 2.
\end{align*}
\]

#### 9.2.2 Weighted Residual Formulation

The domain of interest presented in Fig. 9.7 has four large triangular elements and five finite volumes. The finite volumes consist of the square of area 2 in the middle and the four triangles in the corners. In other words, there is one square volume element and four small triangular volume elements.

We begin by employing Eq. (9.26). We have for our problem
Figure 9.7: Definition drawing of finite volume problem presented as ‘example one’ in Section 9.2.

\[
\begin{align*}
\sum_{e=1}^{E_i} \sum_{j=1}^{5} u_j \int_{\partial \Omega_e} \left( 2 \frac{\partial \phi_j (x, y)}{\partial x} n_x + \frac{\partial \phi_j (x, y)}{\partial y} n_y \right) ds \\
+ \sum_{e=1}^{E_i} \sum_{j=1}^{5} u_j \int_{\Omega_e} (\kappa \phi_j^2 (x, y)) w_i (x, y) d\Omega - \int_{\Omega} Q w_i (x, y) d\Omega + \\
\int_{\partial \Omega} \left[ 2 \frac{\partial \tilde{u} (x, y)}{\partial x} n_x + \frac{\partial \tilde{u} (x, y)}{\partial y} n_y \right] w_i (x, y) ds = 0 \quad i = 1, 2, \ldots, 5
\end{align*}
\]  

(9.32)

where the number of elements identified with (attached to) each node \( i \), where \( i \) identifies the finite volume of interest, is given by \( E_i \). If we write Eq. (9.32) in matrix form we have

\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} & 0 \\
  a_{21} & a_{22} & a_{23} & 0 & a_{25} \\
  a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
  a_{41} & 0 & a_{43} & a_{44} & a_{45} \\
  0 & a_{52} & a_{53} & a_{54} & a_{55}
\end{bmatrix}
\begin{bmatrix}
  u_1 \\
  u_2 \\
  u_3 \\
  u_4 \\
  u_5
\end{bmatrix}
= \begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3 \\
  f_4 \\
  f_5
\end{bmatrix}
\]  

(9.33)

The number of rows in the coefficient matrix in Eq. (9.33) is five, which corresponds to the number of finite volumes. The number of columns is also five and represents the number of unknown values of \( u_j \) at the nodes.

At this point we realize that we actually have only one unknown, namely the value of \( u_j \) at node 3, because the values of \( u_j \) at other nodes are all provided by the boundary conditions. Thus, due to the imposition of Dirichlet boundary conditions,
the rows associated with these known values can be eliminated. As a result this system of equations reduces to

\[
\begin{bmatrix}
    a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
    1 & 0 & 1 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
    u_3 \\
    1
\end{bmatrix}
= \begin{bmatrix}
    f_3
\end{bmatrix}
\] (9.34)

or

\[
(a_{33}) (u_3) = f_3 - (a_{31}) (1) - (a_{34}) (1).
\] (9.35)

### 9.2.3 Element Coefficient Matrices

The first step in evaluating the coefficients is to recognize that \( e = 1, 2, 3, 4 \) and therefore we have four \( 3 \times 3 \) element coefficient matrices of the form.

\[
A = \begin{bmatrix}
    a_{11} & a_{12} & a_{13} \\
    a_{21} & a_{22} & a_{23} \\
    a_{31} & a_{32} & a_{33}
\end{bmatrix}
\] (9.36)

where \( I \) are local coordinates (the concept of local coordinates for triangles was introduced in Section 8.2.1 on page 196). Recall once again that the rows represent the nodes defining where the finite volumes are centered and the columns with the nodes that define the location of the triangular basis functions.

The local coefficient matrix for element \( e \) is (where we again assume \( \phi_I(x, y) \) can be represented by \( \phi_I \))

\[
\begin{bmatrix}
    \int_{\Omega^I_e} \left( 2 \frac{\partial \phi_I}{\partial x} n_x + \frac{\partial \phi_I}{\partial y} n_y \right) ds \\
    + \int_{\Omega^I_e} (\kappa \phi_I) w_I d\Omega \\
    \int_{\Omega^I_{e+1}} \left( 2 \frac{\partial \phi_I}{\partial x} n_x + \frac{\partial \phi_I}{\partial y} n_y \right) ds \\
    + \int_{\Omega^I_{e+1}} (\kappa \phi_I) w_{I+1} d\Omega \\
    \int_{\Omega^I_{e+2}} \left( 2 \frac{\partial \phi_I}{\partial x} n_x + \frac{\partial \phi_I}{\partial y} n_y \right) ds \\
    + \int_{\Omega^I_{e+2}} (\kappa \phi_I) w_{I+2} d\Omega
\end{bmatrix}
\]

\[
\begin{bmatrix}
    \int_{\Omega^I_e} \left( 2 \frac{\partial \phi_{I+1}}{\partial x} n_x + \frac{\partial \phi_{I+1}}{\partial y} (x, y) n_y \right) ds \\
    + \int_{\Omega^I_e} (\kappa \phi_{I+1}) w_I d\Omega \\
    \int_{\Omega^I_{e+1}} \left( 2 \frac{\partial \phi_{I+1}}{\partial x} n_x + \frac{\partial \phi_{I+1}}{\partial y} n_y \right) ds \\
    + \int_{\Omega^I_{e+1}} (\kappa \phi_{I+1}) w_{I+1} d\Omega \\
    \int_{\Omega^I_{e+2}} \left( 2 \frac{\partial \phi_{I+1}}{\partial x} n_x + \frac{\partial \phi_{I+1}}{\partial y} n_y \right) ds \\
    + \int_{\Omega^I_{e+2}} (\kappa \phi_{I+1}) w_{I+2} d\Omega
\end{bmatrix}.
\] (9.37)

The table relating the global and local numbering for the problem shown in Fig. 9.7:

If the values in the above table are introduced into Eq. (9.37) one obtains for the global coefficient matrix elements
9.2. Finite Volume Example Problem

Global

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>Element</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>1</td>
<td></td>
<td>I+1</td>
<td>I+2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I+2</td>
<td>I+1</td>
<td>I+1</td>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I+2</td>
<td>I+1</td>
<td>I</td>
<td>I+2</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>I+1</td>
<td>I</td>
<td>I+1</td>
<td>I+2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Table 9.1: Table relating global and local coordinates.

\[ a_{31} = \sum_{e=1}^{E_3} \int_{\partial \Omega_3^e} \left( 2 \frac{\partial \phi_1 (x, y)}{\partial x} n_x + \frac{\partial \phi_1 (x, y)}{\partial y} n_y \right) ds + \int_{\Omega_3} (\kappa \phi_1 (x, y)) w_3 (x, y) d\Omega \] (9.38)

\[ a_{33} = \sum_{e=1}^{E_3} \int_{\partial \Omega_3^e} \left( 2 \frac{\partial \phi_3 (x, y)}{\partial x} n_x + \frac{\partial \phi_3 (x, y)}{\partial y} n_y \right) ds + \int_{\Omega_3} (\kappa \phi_3 (x, y)) w_3 (x, y) d\Omega \] (9.39)

\[ a_{34} = \sum_{e=1}^{E_3} \int_{\partial \Omega_3^e} \left( 2 \frac{\partial \phi_4 (x, y)}{\partial x} n_x + \frac{\partial \phi_4 (x, y)}{\partial y} n_y \right) ds + \int_{\Omega_3} (\kappa \phi_4 (x, y)) w_3 (x, y) d\Omega \] (9.40)

\[ f_3 = \int_{\Omega_3} Qw_i (x, y) d\Omega \] (9.41)

where the subscripts on the \( \phi_i (x, y) \) values are in global coordinates.

9.2.4 Evaluation of the Line Integral

The global coefficient \( a_{31} \) contributed by element 1 is calculated for basis function \( \phi_1 \) and weighting function \( w_3 \) as follows. First we substitute the values of \( I = 3 \) and \( J = 1 \) to give

\[ \int_{\partial \Omega_3^e} \left( 2 \frac{\partial \phi_1 (x, y)}{\partial x} n_x + \frac{\partial \phi_1 (x, y)}{\partial y} n_y \right) ds \]

\[ = \int_{DE} \left( 2 \frac{\partial \phi_1 (x, y)}{\partial x} n_x \big|_{DE} + \frac{\partial \phi_1 (x, y)}{\partial y} n_y \big|_{DE} \right) ds + \int_{EA} \left( 2 \frac{\partial \phi_1 (x, y)}{\partial x} n_x \big|_{EA} + \frac{\partial \phi_1 (x, y)}{\partial y} n_y \big|_{EA} \right) ds \] (9.42)

where \( |_{DE} \) is the segment of side \( \partial \Omega_3^e \) that is composed of the line \( \overline{EA} \) and \( |_{DE} \) is the segment along the line \( \overline{DE} \). Now we recall the definition of the the basis and obtain their derivatives, that is

\[ \phi_1 (x, y) = \frac{x(y_{I+1} - y_{I+2}) + y(x_{I+2} - x_{I+1}) + (x_{I+1}y_{I+2} - x_{I+2}y_{I+1})}{\det [P]} \] (9.43)
from which

$$\frac{\partial \phi_I}{\partial x} = \frac{(y_{I+1} - y_{I+2})}{\det [P]} \quad (9.44)$$

and

$$\frac{\partial \phi_I}{\partial y} = \frac{(x_{I+2} - x_{I+1})}{\det [P]} \quad (9.45)$$

Therefore, Eq. (9.42) becomes

$$\int_{DE} 2(y_3 - y_4) \frac{n_x|_{DE}}{P} + (x_4 - x_3) \frac{n_y|_{DE}}{P} \, ds \quad (9.46)$$

$$+ \int_{AE} 2(y_3 - y_4) \frac{n_x|_{AE}}{P} + (x_4 - x_3) \frac{n_y|_{AE}}{P} \, ds.$$

Next observe in Fig. 9.7, the location of the line that defines the side of the triangle connecting the location of node 3, which identifies the volume of interest, and the node 1 which identifies the element basis function of interest. If we now imagine being at the node 3 and direct our attention towards node 1, we observe that the resulting line is given by \((x_3 - x_1, y_3 - y_1)\). We now need the \(n_x\) and \(n_y\) values to complete the evaluation associated with Eq. (9.46). From Eq. (9.22) we see that the value of \(n_x\) is given by

$$n_x|_{DE} = \frac{x_4 - x_3}{\sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2}}$$

$$= \frac{0 - 1}{\sqrt{(1)^2 + (1)^2}}$$

$$= -\frac{1}{\sqrt{2}}. \quad (9.47)$$

Similarly for the evaluation of \(n_y\), we have

$$n_y|_{DE} = \frac{y_4 - y_3}{\sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2}}$$

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

$$= \frac{1}{\sqrt{2}}. \quad (9.48)$$

Now we can combine this information to give

$$n|_{DE} = (n_x|_{DE}, n_y|_{DE})$$

$$= \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right). \quad (9.49)$$
But we still need to have the information relevant to side $EA$. Following the strategy presented above we get

\[
n_{x|EA} = \frac{x_1 - x_3}{\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} = \frac{0 - 1}{\sqrt{1^2 + 1^2}} = -\frac{1}{\sqrt{2}}.
\]

(9.50)

and

\[
n_{y|EA} = \frac{y_1 - y_3}{\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} = \frac{0 - 1}{\sqrt{1^2 + 1^2}} = -\frac{1}{\sqrt{2}}.
\]

(9.51)

such that

\[
n_{|EA} = (n_{x|EA}, n_{y|EA}) = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right).
\]

(9.52)

We now substitute these values into Eq. (9.42) to give

\[
\int_{DE} \frac{2(y_3 - y_4) n_{x|DE} + (x_4 - x_3) n_{y|DE}}{\det [P]} ds
\]

\[
+ \int_{EA} \frac{2(y_3 - y_4) n_{x|EA} + (x_4 - x_3) n_{y|EA}}{\det [P]} ds
\]

\[
= \int_{DE} \frac{2(y_3 - y_4) - \frac{1}{\sqrt{2}} + (x_4 - x_3) \frac{1}{\sqrt{2}}}{\det [P]} ds
\]

\[
+ \int_{EA} \frac{2(y_3 - y_4) - \frac{1}{\sqrt{2}} + (x_4 - x_3) - \frac{1}{\sqrt{2}}}{\det [P]} ds.
\]

(9.53)
After substituting the coordinate locations of \( x_i, y_i \) we get

\[
\int_{DE} \frac{2(y_3 - y_4) \frac{1}{\sqrt{2}} + (x_4 - x_3) \frac{1}{\sqrt{2}}}{\text{det}[P]} ds \\
+ \int_{EA} \frac{2(y_3 - y_4) \frac{1}{\sqrt{2}} + (x_4 - x_3) \frac{1}{\sqrt{2}}}{\text{det}[P]} ds \\
= \int_{DE} \frac{2(1 - 2) \frac{1}{\sqrt{2}} + (0 - 1) \frac{1}{\sqrt{2}}}{\text{det}[P]} ds \\
+ \int_{EA} \frac{2(1 - 2) \frac{1}{\sqrt{2}} + (0 - 1) \frac{1}{\sqrt{2}}}{\text{det}[P]} ds \\
= \int_{DE} \frac{2 - 1}{\sqrt{2} \text{det}[P]} ds \\
+ \int_{EA} \frac{2 + 1}{\sqrt{2} \text{det}[P]} ds.
\]

(9.54)

We can calculate the \( \text{det}[P] \) from the coordinate matrix which, for this element, is defined using the above table as

\[
[P] = \begin{bmatrix}
  x_1 & y_1 & 1 \\
  x_{I+1} & y_{I+1} & 1 \\
  x_{I+2} & y_{I+2} & 1
\end{bmatrix} = \begin{bmatrix}
  x_1 & y_1 & 1 \\
  x_3 & y_3 & 1 \\
  x_4 & y_4 & 1
\end{bmatrix} = \begin{bmatrix}
  0 & 0 & 1 \\
  1.4 & 0 & 1 \\
  1.4 & 1.4 & 1
\end{bmatrix}.
\]

(9.55)

We can determine \( \text{det}[P] \) either by calculation from Eq. (9.55), or from the known relationship that \( \text{det}[P] = 2A \) presented earlier in Section 8.2.1 on page 196. In either event we determine that \( \text{det}[P] = 2 \) and Eq. (9.54) becomes

\[
\int_{DE} \frac{2 - 1}{2\sqrt{2}} ds + \int_{EA} \frac{2 + 1}{2\sqrt{2}} ds \\
\int_{DE} \frac{1}{2\sqrt{2}} ds + \int_{EA} \frac{2 + 1}{2\sqrt{2}} ds \\
= \left( \frac{1}{2\sqrt{2}} \right) |DE| + \left( \frac{2 + 1}{2\sqrt{2}} \right) |EA| \\
= \left( \frac{1}{2\sqrt{2}} \right) \sqrt{2} + \left( \frac{2 + 1}{2\sqrt{2}} \right) \sqrt{2} \\
= \frac{4}{4} = 1.
\]

(9.56)

Thus the line integral component of the coefficient for \( I + 1 = 3, J = 1 \) for element 1 is 1.0.

As a second example consider \( I + 1 = 3, J + 2 = 1 \) and substitute this information
in Eq. (9.42) for element 2. We obtain
\[
\int_{\partial \Omega_2^3} \left( 2 \frac{\partial \phi_1}{\partial x} n_x + \frac{\partial \phi_1}{\partial y} n_y \right) ds \\
= \int_{AF} \left( 2 \frac{\partial \phi_1}{\partial x} n_x |_{AF} + \frac{\partial \phi_1}{\partial y} n_y |_{AF} \right) ds \\
+ \int_{FB} \left( 2 \frac{\partial \phi_1}{\partial x} n_x |_{FB} + \frac{\partial \phi_1}{\partial y} n_y |_{FB} \right) ds \\
= \left( 2 \frac{\partial \phi_1}{\partial x} n_x |_{AF} + \frac{\partial \phi_1}{\partial y} n_y |_{AF} \right) |AF| \\
+ \left( 2 \frac{\partial \phi_1}{\partial x} n_x |_{FB} + \frac{\partial \phi_1}{\partial y} n_y |_{FB} \right) |FB| \\
= \left[ 2 \left( -\frac{1}{2} \right) \left( -\frac{\sqrt{2}}{2} \right) + \left( -\frac{1}{2} \right) \left( -\frac{\sqrt{3}}{2} \right) \right] \frac{\sqrt{2}}{2} \\
+ \left[ 2 \left( -\frac{1}{2} \right) \left( \frac{\sqrt{2}}{2} \right) + \left( -\frac{1}{2} \right) \left( -\frac{\sqrt{2}}{2} \right) \right] \frac{\sqrt{2}}{2} \\
= \left[ \left( \frac{3}{4} \right) \sqrt{2} - \frac{1}{4} \sqrt{2} \right] \frac{\sqrt{2}}{2} \\
= 0.5 
\] (9.57)
so the coefficient for \( I + 1 = 3, J + 2 = 2 \) is 0.5 in element 2.

In this, the third example, we consider the corner finite volume, that is the one associated with node 2, element 3, and the basis function identified with node 2, that is \( I + 2 = 2 \) and \( J + 2 = 2 \). Substitution in Eq. (9.42) gives
\[
\int_{\partial \Omega_2^3} \left( 2 \frac{\partial \phi_2}{\partial x} n_x + \frac{\partial \phi_2}{\partial y} n_y \right) ds. 
\] (9.58)
The finite volume perimeter of interest is
\[
\partial \Omega_2^3 = \overline{2G} + \overline{GB}. 
\] (9.59)
so Eq. (9.58) becomes, where \( \overline{2G} \) is the segment connecting node 2 and point \( G \),
\[
\int_{\partial \Omega_2^3} \left( 2 \frac{\partial \phi_2}{\partial x} n_x + \frac{\partial \phi_2}{\partial y} n_y \right) ds \\
= \int_{\overline{2G}} \left( 2 \frac{\partial \phi_2}{\partial x} n_x + \frac{\partial \phi_2}{\partial y} n_y \right) ds \\
+ \int_{\overline{GB}} \left( 2 \frac{\partial \phi_2}{\partial x} n_x + \frac{\partial \phi_2}{\partial y} n_y \right) ds. 
\] (9.60)
The normal vector to \( \overline{2G} \) is given by (see Eqs. (9.47) to (9.49))
\[
\mathbf{n}|_{\overline{2G}} = (n_x|_{\overline{2G}}, n_y|_{\overline{2G}}) \\
= (0, 1) 
\] (9.62)
and
\[
\mathbf{n}_{|_{\Omega_B}} = (n_x|_{\Omega_B}, n_y|_{\Omega_B}) = \left( -\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right). \tag{9.63}
\]

The next step is to obtain the derivatives of the basis function for \( \phi_2(x, y) \) which we obtain from Eqs. (9.44) and (9.45), that is
\[
\frac{\partial \phi_2}{\partial x} = \frac{y_5 - y_3}{\det [P]} = \frac{1}{2} \tag{9.64}
\]
and
\[
\frac{\partial \phi_2}{\partial y} = \frac{x_3 - x_5}{\det [P]} = -\frac{1}{2}. \tag{9.65}
\]

We can now evaluate the equation of interest, namely
\[
\begin{align*}
\int_{2G} & \left( 2 \frac{\partial \phi_2}{\partial x} n_x + \frac{\partial \phi_2}{\partial y} n_y \right) ds \\
& + \int_{\Omega_B} \left( 2 \frac{\partial \phi_2}{\partial x} n_x + \frac{\partial \phi_2}{\partial y} n_y \right) ds \\
& = \left( 2 \frac{\partial \phi_2}{\partial x} n_x|_{2G} + \frac{\partial \phi_2}{\partial y} n_y|_{2G} \right) |2G| \\
& + \left( 2 \frac{\partial \phi_2}{\partial x} n_x|_{\Omega_B} + \frac{\partial \phi_2}{\partial y} n_y|_{\Omega_B} \right) |\Omega_B| \\
& = \left[ 2 \left( \frac{1}{2} \right) \right] (1) + \left[ -\frac{1}{2} \right] (0) \tag{1} \\
& + \left[ 2 \left( \frac{1}{2} \right) \left( -\frac{\sqrt{2}}{2} \right) + \left( -\frac{1}{2} \right) \left( \frac{\sqrt{2}}{2} \right) \right] \left( \frac{\sqrt{2}}{2} \right) \\
& = 1 - \left( \frac{3\sqrt{2}}{4} \right) \left( \frac{\sqrt{2}}{2} \right) \\
& = 0.25. \tag{9.66}
\end{align*}
\]

Thus, the coefficient located at \( I + 2 = 2, J + 2 = 2 \) in element 3 is 0.25.

Finally, we consider the case of the contribution to coefficient \( a_{33} \) attributable to element 3, that is \( I + 1 = 3 \) and \( J + 1 = 3 \) The two segments of the boundary of the finite volume of interest are
\[
\partial \Omega_3^3 = BG + GC. \tag{9.67}
\]

The term of interest is
9.2. FINITE VOLUME EXAMPLE PROBLEM 1

\[ \int_{\partial \Omega_3} \left( 2 \frac{\partial \phi_3}{\partial x} n_x + \frac{\partial \phi_3}{\partial y} n_y \right) ds \]

\[ = \int_{BG} \left( 2 \frac{\partial \phi_3}{\partial x} n_x + \frac{\partial \phi_3}{\partial y} n_y \right) ds \]

\[ + \int_{GC} \left( 2 \frac{\partial \phi_3}{\partial x} n_x + \frac{\partial \phi_3}{\partial y} n_y \right) ds \]

\[ = \left( 2 \frac{\partial \phi_3}{\partial x} n_x|_{BG} + \frac{\partial \phi_3}{\partial y} n_y|_{BG} \right) |BG| \]

\[ + \left( 2 \frac{\partial \phi_3}{\partial x} n_x|_{GC} + \frac{\partial \phi_3}{\partial y} n_y|_{GC} \right) |GC|. \] (9.68)

The derivatives of the basis function \( \phi_3(x, y) \) are

\[ \frac{\partial \phi_3}{\partial x} = \frac{y_2 - y_5}{\det [P]} \] (9.69)

and

\[ \frac{\partial \phi_3}{\partial x} = \frac{x_5 - x_2}{\det [P]}. \] (9.70)

The values of \( n_x|_{BG} \) and \( n_x|_{GC} \) are computed as follows:

\[ n|_{BG} = (n_x|_{BG}, n_y|_{GC}) \]

\[ = \frac{1}{\sqrt{(x_2 - x_3)^2 + (y_2 - y_3)^2}} ((x_2 - x_3), (y_2 - y_3)) \]

\[ = \left( \frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2} \right) \] (9.71)

\[ n|_{GC} = (n_x|_{GC}, n_y|_{GC}) \]

\[ = \frac{1}{\sqrt{(x_5 - x_3)^2 + (y_5 - y_3)^2}} ((x_5 - x_3), (y_5 - y_3)) \]

\[ = \left( \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right) . \] (9.72)

We can now evaluate our coefficient

\[ \left( 2 \frac{\partial \phi_3}{\partial x} n_x|_{BG} + \frac{\partial \phi_3}{\partial y} n_y|_{BG} \right) |BG| \]

\[ + \left( 2 \frac{\partial \phi_3}{\partial x} n_x|_{GC} + \frac{\partial \phi_3}{\partial y} n_y|_{GC} \right) |GC| \]

\[ = 2 (-1) \left( \frac{\sqrt{2}}{2} \right) \frac{\sqrt{2}}{2} + \left[ 2 (-1) \left( \frac{\sqrt{2}}{2} \right) \frac{\sqrt{2}}{2} \right] \]

\[ = -1 - 1 = -2. \] (9.73)
Thus the coefficient \( i = 3, \ j = 3 \) of element 3 is \(-2\).

If we evaluate all the coefficients in each element coefficient matrix we obtain for the line-integral component:

\[
\text{Element 1} \\
\begin{bmatrix}
0.25 & -1.0 & 0.75 \\
1.0 & -2.0 & 1.0 \\
0.75 & -1.0 & 0.25
\end{bmatrix}
\] (9.74)

\[
\text{Element 2} \\
\begin{bmatrix}
-0.25 & -0.50 & 0.75 \\
0.50 & -1.0 & 0.50 \\
0.75 & -0.50 & -0.25
\end{bmatrix}
\] (9.75)

\[
\text{Element 3} \\
\begin{bmatrix}
0.25 & -1.0 & 0.75 \\
1.0 & -2.0 & 1.0 \\
0.75 & -1.0 & 0.25
\end{bmatrix}
\] (9.76)

\[
\text{Element 4} \\
\begin{bmatrix}
-0.25 & -0.50 & 0.75 \\
0.50 & -1.0 & 0.50 \\
0.75 & -0.50 & -0.25
\end{bmatrix}
\] (9.77)

### 9.2.5 Evaluation of the Area Integral

We need to also evaluate integrals in Eq. (9.32) of the form

\[
\int_{\Omega} (\kappa \phi_j (x,y) \ d\Omega) \ w_i (x,y) \ d\Omega = \int_{\Omega'} \kappa \phi_j (x,y) \ d\Omega
\]

for the system shown in Fig. 9.7. As an example of how this can be done consider element 1 shown in Fig. 9.8 which is the same as element 1 in Fig. 9.7 rotated clockwise 45 degrees so that the sides of the element line up with the \( x \) and \( y \) axes. The value of the integral does not change by doing this, but the presentation is simplified. Let values of \( i \) and \( j \) both be 3.

The finite volume of interest is the shaded square area in Fig. 9.8. In this example, the length of the line \( BC \) is one half the length of the side \( CE \). The basis function in element 1 identified with node 3 has a value of 1 at node 3 and zero at nodes 1 and 4. Thus, at the point \( c \) the value of \( \phi_3 \) is 1.0, at the points \( b \) and \( d \) the values are 0.5, and at \( e \) and \( g \) the values are 0.

Two approaches to obtaining this integral are presented. The first is based upon geometric considerations and the second upon numerical integration.
9.2. FINE ME ME EXAMPLE PROBLEM 1

To obtain the integral of \( \phi_3 \) located under the shaded square, one first calculates the area under the triangle \( \Delta \text{ecg} \). This is obtained from the equation given in Eq. (8.83) and the integration formula 8.121. Using this information, the formula for evaluating the integral is

\[
\int_{\Omega_{\text{ecg}}} \phi_3 \, d\omega = \frac{\det [P]_{\text{ecg}}/2}{3} \quad (9.78)
\]

where \( \det [P]_{\text{ecg}} \) is the determinant of the matrix \( [P] \) defined for triangle \( \text{ecg} \).

Next subtract the integral

\[
\int_{\Omega_{\text{eba}}} \phi_3 \, d\omega = \frac{\det [P]_{\text{eba}}/2}{6} = \frac{1}{24} \quad (9.79)
\]

and the integral

\[
\int_{\Omega_{\text{adg}}} \phi_3 \, d\omega = \frac{\det [P]_{\text{adg}}/2}{6} = \frac{1}{24} \quad (9.80)
\]

from Eq. (9.78). The result is the portion of the integral in Eq. (9.78) under the weighting function associated with the finite volume located at 3. This calculation provides the element coefficient value at \( i = 3, j = 3 \). More specifically, the coefficient value is

\[
a_{33} = \int_{\Omega^3} \phi_3 (x, y) \, d\omega = \frac{(\det \Delta \text{ecg})/2}{3} - \frac{(\det \Delta \text{eba})/2}{6} - \frac{(\det \Delta \text{adg})/2}{6} = \frac{1}{3} - \frac{0.25}{6} - \frac{0.25}{6} = \frac{1}{4} \quad (9.81)
\]

To obtain the value for coefficient at \( i = 3, j = 4 \), one uses the value obtained in Eq. (9.79) and for coefficient at \( i = 3, j = 1 \) the value is given by Eq. (9.80). Evaluation of all of the integrals of the form \( \int_{\Omega^i} \phi_j (x, y) \, d\omega \) for element 1 gives the coefficient matrix:

Figure 9.8: Representation of the finite element configuration for calculation of the area integral.

**Geometric Approach**

To obtain the integral of \( \phi_3 \) located under the shaded square, one first calculates the area under the triangle \( \Delta \text{ecg} \). This is obtained from the equation given in Eq. (8.83) and the integration formula 8.121. Using this information, the formula for evaluating the integral is

\[
\int_{\Omega_{\text{ecg}}} \phi_3 \, d\omega = \frac{\det [P]_{\text{ecg}}/2}{3} \quad (9.78)
\]

where \( \det [P]_{\text{ecg}} \) is the determinant of the matrix \( [P] \) defined for triangle \( \text{ecg} \).

Next subtract the integral

\[
\int_{\Omega_{\text{eba}}} \phi_3 \, d\omega = \frac{\det [P]_{\text{eba}}/2}{6} = \frac{1}{24} \quad (9.79)
\]

and the integral

\[
\int_{\Omega_{\text{adg}}} \phi_3 \, d\omega = \frac{\det [P]_{\text{adg}}/2}{6} = \frac{1}{24} \quad (9.80)
\]

from Eq. (9.78). The result is the portion of the integral in Eq. (9.78) under the weighting function associated with the finite volume located at 3. This calculation provides the element coefficient value at \( i = 3, j = 3 \). More specifically, the coefficient value is

\[
a_{33} = \int_{\Omega^3} \phi_3 (x, y) \, d\omega = \frac{(\det \Delta \text{ecg})/2}{3} - \frac{(\det \Delta \text{eba})/2}{6} - \frac{(\det \Delta \text{adg})/2}{6} = \frac{1}{3} - \frac{0.25}{6} - \frac{0.25}{6} = \frac{1}{4} \quad (9.81)
\]

To obtain the value for coefficient at \( i = 3, j = 4 \), one uses the value obtained in Eq. (9.79) and for coefficient at \( i = 3, j = 1 \) the value is given by Eq. (9.80). Evaluation of all of the integrals of the form \( \int_{\Omega^i} \phi_j (x, y) \, d\omega \) for element 1 gives the coefficient matrix:
Because of the geometry, the coefficient matrix for each element, in terms of the local coordinates $I, I + 1,$ and $I + 2,$ is the same in this example.

### Numerical Integration Approach

While the above strategy works for the topology of the special case we are considering, it is awkward to use in a more general case. An alternative approach uses the isoparametric transformation concept which we introduced in Section 8.3. In this approach we transform our coordinate system from $(x, y)$ to $(\xi, \eta)$ where $-1 \leq \xi, \eta \leq 1.$ The first step is to formulate the transformation equations. As earlier we proceed as follows:

$$ x = \sum_{i=1}^{4} x_i \psi_i (\xi, \eta) \equiv P (\xi, \eta) \quad (9.83) $$

and

$$ y = \sum_{i=1}^{4} y_i \psi_i (\xi, \eta) \equiv Q (\xi, \eta) \quad (9.84) $$

where, referring to Fig. 8.17 we have

$$ \psi_1 (\xi, \eta) = \frac{1}{4} (1 - \xi) (1 - \eta) \quad (9.85) $$

$$ \psi_2 (\xi, \eta) = \frac{1}{4} (1 + \xi) (1 - \eta) \quad (9.86) $$

$$ \psi_3 (\xi, \eta) = \frac{1}{4} (1 + \xi) (1 + \eta) \quad (9.87) $$

$$ \psi_4 (\xi, \eta) = \frac{1}{4} (1 - \xi) (1 + \eta). \quad (9.88) $$

For the problem at hand we need to transform our integral from the $(x, y)$ system to the $(\xi, \eta)$ system. The new integral takes the form

$$ \int_{\Omega} \phi_i (x, y) dx = \int_{\xi=-1}^{\xi=1} \int_{\eta=-1}^{\eta=1} \phi_i [P (\xi, \eta), Q (\xi, \eta)] |J (\xi, \eta)| d\xi d\eta \quad (9.89) $$

where $J (\xi, \eta)$ is the Jacobian of the transformation and defined as

$$ \det [J (\xi, \eta)] = \det \begin{vmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{vmatrix}. \quad (9.90) $$

In our example, the square area of interest is such that the $(x, y)$ coordinates are colinear with $(\xi, \eta).$ This simplifies the problem since some of the terms in the transformation vanish and we get for $\frac{\partial x}{\partial \xi}$
\[
\frac{\partial x}{\partial \xi} = x_1 \frac{\partial}{\partial \xi} \psi_1 (\xi, \eta) + x_2 \frac{\partial}{\partial \xi} \psi_2 (\xi, \eta) + x_3 \frac{\partial}{\partial \xi} \psi_3 (\xi, \eta) + x_4 \frac{\partial}{\partial \xi} \psi_4 (\xi, \eta)
\]
\[
= \frac{1}{4} \left\{ (0) - (1 - \eta) + (0.35) (1 - \eta) + (0.35) (1 + \eta) + (0) - (1 + \eta) \right\}
\]
\[
= \frac{1}{4} \left\{ (0.35) (1 - \eta) + (0.35) (1 + \eta) \right\}
\]
\[
= \frac{0.7}{2} = 0.35
\]
(9.91)

and for \( \frac{\partial y}{\partial \xi} \)
\[
\frac{\partial y}{\partial \xi} = y_1 \frac{\partial}{\partial \xi} \psi_1 (\xi, \eta) + y_2 \frac{\partial}{\partial \xi} \psi_2 (\xi, \eta) + y_3 \frac{\partial}{\partial \xi} \psi_3 (\xi, \eta) + y_4 \frac{\partial}{\partial \xi} \psi_4 (\xi, \eta)
\]
\[
= \frac{1}{4} \left\{ (0) - (1 - \eta) + (0) (1 - \eta) + (0.35) (1 + \eta) + (0.35) [ (1 - \eta) ] \right\}
\]
\[
= \frac{1}{4} (0.7) (1 + \eta) - (0.7) (1 + \eta)
\]
\[
= 0.
\]
(9.92)

Using a similar strategy for the other elements of the Jacobian we get
\[
\text{det} [J] = \text{det} \begin{bmatrix} 0.35 & 0 \\ 0 & 0.35 \end{bmatrix} = 0.125.
\]
(9.93)

The integral appearing in Eq. (9.89) now takes the form
\[
\int_{\Omega_i} \phi_i (x, y) \, dx = \int_{\xi=-1}^{\xi=1} \int_{\eta=-1}^{\eta=1} \phi_i [P (\xi, \eta) , Q (\xi, \eta)] \, |J (\xi, \eta)| \, d\xi d\eta
\]
\[
= \int_{\xi=-1}^{\xi=1} \int_{\eta=-1}^{\eta=1} \phi_i [P (\xi, \eta) , Q (\xi, \eta)] \, (0.12) \, d\xi d\eta.
\]
(9.94)

So, for the case of \( \phi_i (x, y) = \phi_3 (x, y) \) we have
\[
\int_{\Omega_i} \phi_3 (x, y) \, dx = \int_{\xi=-1}^{\xi=1} \int_{\eta=-1}^{\eta=1} \phi_3 [P (\xi, \eta) , Q (\xi, \eta)] \, (0.12) \, d\xi d\eta.
\]
(9.95)

The next step is to determine \( \phi_3 [P (\xi, \eta) , Q (\xi, \eta)] \). To do this we need to write out the definition of \( \phi_3 (x, y) \) which is obtained from the definition we have used earlier, namely
\[
\phi_I (x, y) = \frac{x (y_{I+1} - y_{I+2}) + y (x_{I+2} - x_{I+1}) + (x_{I+1} y_{I+2} - x_{I+2} y_{I+1})}{\det [P]}.
\]
(9.96)

Substitution for the values of \( I, I + 2, \) and \( I + 3 \) from Fig. 9.8 into this expression we get
\[
\phi_3 [P (\xi, \eta) , Q (\xi, \eta)] = \frac{(P) (y_4 - y_1) + (Q) (x_1 - x_4) + (x_4 y_1 - x_1 y_4)}{\det [P]}
\]
(9.97)

which, after providing the location values, gives
\[
\phi_3 (\xi, \eta) = \frac{(P)(\sqrt{2} - 0) + (Q)(0 - \sqrt{2}) + ((\sqrt{2})(0) - (0)(\sqrt{2}))}{\det [P]} \tag{9.98}
\]

or
\[
\phi_3 (\xi, \eta) = \frac{(P)(\sqrt{2}) - (Q)(\sqrt{2})}{2} \tag{9.99}
\]

\[
\int_{\Omega_3^n} \phi_3 (x, y) \, dx = \int_{\xi = -1}^{\xi = 1} \int_{\eta = -1}^{\eta = 1} \phi_3 (P(\xi, \eta), Q(\xi, \eta)) \, (0.12) \, d\xi d\eta
\]

\[
= \int_{\xi = -1}^{\xi = 1} \int_{\eta = -1}^{\eta = 1} \frac{(P)(\sqrt{2}) - (Q)(\sqrt{2})}{2} \, (0.12) \, d\xi d\eta
\]

\[
= \int_{\xi = -1}^{\xi = 1} \int_{\eta = -1}^{\eta = 1} \left( \frac{\sqrt{2}}{2} \right) \left( \sum_{i=1}^{i=4} x_i \psi_i (\xi, \eta) - \sum_{i=1}^{i=4} y_i \psi_i (\xi, \eta) \right) \, (0.12) \, d\xi d\eta
\]

This looks pretty messy, but there is an easy solution to the evaluation of this expression. Recall that we introduced Gaussian quadrature in Section 3.1.6. Let us try one Gauss point to see how well this procedure approximates this integral. The value of the single Gauss point is \((0, 0)\) and the weighting is 2. Using this information, Eq. (9.100) becomes

\[
\int_{\Omega_3^n} \phi_3 (x, y) \, dx \approx \int_{\xi = -1}^{\xi = 1} \int_{\eta = -1}^{\eta = 1} \frac{\sqrt{2}}{2} \left( \sum_{i=1}^{i=4} x_i \psi_i (\xi, \eta) - \sum_{i=1}^{i=4} y_i \psi_i (\xi, \eta) \right) \, (0.12) \, d\xi d\eta
\]

\[
= 2\sqrt{2} \left[ x_1 \psi_1 (0, 0) + x_2 \psi_2 (0, 0) + x_3 \psi_3 (0, 0) + x_4 \psi_4 (0, 0) - y_1 \psi_1 (0, 0) - y_2 \psi_2 (0, 0) - y_3 \psi_3 (0, 0) - y_4 \psi_4 (0, 0) \right] (0.12)
\]

\[
= 2\sqrt{2} \left[ \left( \frac{\sqrt{2}}{2} \right) (0.25) + \left( \frac{\sqrt{2}}{2} \right) (0.25) + \left( \frac{\sqrt{2}}{2} \right) (0.25) + \left( \frac{\sqrt{2}}{2} \right) 0.25 \right. \\
\left. - (0) (0.25) - (0) (0.25) - \left( \frac{\sqrt{2}}{2} \right) (0.25) - \left( \frac{\sqrt{2}}{2} \right) (0.25) \right] (0.12)
\]

\[
= 0.25. \tag{9.101}
\]

This is the same value we obtained from geometric consideration in Eq. (9.81). Note that this approach is general and holds for any finite volume mesh.

### 9.2.6 Global Matrix Assembly

We are now in a position to evaluate our global coefficients \(a_{31}, a_{33}\) and \(a_{34}\). We do this using the element matrices presented earlier and Table 9.1 relating the global and local matrix coefficients. For example, the line integral global values for \(a_{31}, a_{33}\) and \(a_{34}\) we obtain are as shown in the following table:

<table>
<thead>
<tr>
<th>element</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a_{31})</td>
<td>1.0</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(a_{33})</td>
<td>-2</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
</tr>
<tr>
<td>(a_{34})</td>
<td>1.0</td>
<td>0.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
For the area integral we get

\[
\begin{array}{c|cccc}
\text{element} & 1 & 2 & 3 & 4 \\
\hline
a_{31} & 0.125 & 0.125 & & \\
a_{33} & 0.25 & 0.25 & 0.25 & 0.25 \\
a_{34} & 0.125 & & & 0.125 \\
\end{array}
\]

The complete line-integral global matrix is

\[
\begin{bmatrix}
0 & 0.75 & -1.5 & 0.75 & 0 \\
0.75 & 0 & -1.5 & 0 & 0.75 \\
1.5 & 1.5 & -6.0 & 1.5 & 1.5 \\
0.75 & 0 & -1.5 & 0 & 0.75 \\
0 & 0.75 & -1.5 & 0.75 & 0 \\
\end{bmatrix}
\]

(9.102)

and the complete area integral global matrix is

\[
\begin{bmatrix}
0.33 & 0.042 & 0.084 & 0.042 & 0 \\
0.042 & 0.33 & 0.084 & 0 & 0.042 \\
0.25 & 0.25 & 1.0 & 0.25 & 0.25 \\
0.042 & 0 & 0.084 & 0.33 & 0.042 \\
0 & 0.042 & 0.084 & 0.042 & 0.33 \\
\end{bmatrix}
\]

(9.103)

The combined final coefficient matrix is

\[
\begin{bmatrix}
0.33 & 0.79 & -1.4 & 0.79 & 0 \\
0.79 & 0.33 & -1.4 & 0 & 0.79 \\
1.7 & 1.7 & -5.0 & 1.7 & 1.7 \\
0.79 & 0 & -1.42 & 0.33 & 0.79 \\
0 & 0.79 & -1.42 & 0.79 & 0.33 \\
\end{bmatrix}
\]

(9.104)

Collecting the information from the above table and assuming \( \kappa = 1 \), we obtain

\[
a_{31} = 1.7
\]

(9.105)

\[
a_{33} = -5
\]

(9.106)

\[
a_{34} = 1.75.
\]

(9.107)

Returning to our equation we have

\[
(a_{33}) (u_3) = f_3 - (a_{31}) (1) - (a_{34}) (1) \\
-5u_3 = 0 - (1.7) (1) - (1.7) (1) \\
u_3 = 0.68.
\]

(9.108)

If \( \kappa \) is taken as 0, the result becomes \( u_3 = 0.5 \).
9.3 Finite Volume Example Problem Two

9.3.1 Problem Definition

Example problem 2 is provided to illustrate the application of the finite volume approach when a cross-derivative is involved. The problem is generated by rotation of the problem presented as example 1 and is shown in Fig. 9.9. The equations in the rotated system exhibit a cross derivative. The solution to both the first and second problems should be the same for the case of $\kappa = 0$:

$$1.5 \frac{\partial^2 u(x, y)}{\partial x^2} - 0.5 \frac{\partial^2 u(x, y)}{\partial x \partial y} + 1.5 \frac{\partial^2 u(x, y)}{\partial y^2} - Q = 0 \quad x, y \in \Omega \quad (9.109)$$

$$u(x, y) = 1 \quad \text{line 1 - 4} \quad (9.110)$$

$$u(x, y) = 0 \quad \text{line 2 - 5} \quad (9.111)$$

$$\frac{\partial u(x, y)}{\partial n} = 0 \quad \text{line 1 - 2} \quad (9.112)$$

$$\frac{\partial u(x, y)}{\partial n} = 0 \quad \text{line 4 - 5}. \quad (9.113)$$

Figure 9.9: Definition diagram for example two of the finite volume methods considered in Section 9.3.

The domain of interest is presented in Fig. 9.9, and as in example problem 1, has four triangular elements and five finite volumes. The finite volumes consist of the square in the middle and the four triangles in the corners (note that each corner volume is actually made up of two smaller elements).

9.3.2 Weighted Residual Formulation

Beginning with Eq. (9.109) we have, after substituting our approximation $\hat{u}(x, y)$ for $u(x, y)$, applying the method of weighted residuals using $w_i(x, y)$ as the weighting
function, and finally, employing Green’s theorem to the second-order term
\[
\sum_{e=1}^{E_i} \sum_{j=1}^{5} u_j \int_{\Omega_i} \left( -1.5 \frac{\partial \phi_j}{\partial x} \frac{\partial w_i}{\partial x} + 0.5 \frac{\partial \phi_j}{\partial y} \frac{\partial w_i}{\partial x} + 0.5 \frac{\partial \phi_j}{\partial x} \frac{\partial w_i}{\partial y} - 1.5 \frac{\partial \phi_j}{\partial y} \frac{\partial w_i}{\partial y} \right) \, dw
\]
\[- \int_{\partial \Omega} Qw_i \, d\Omega + \int_{\partial \Omega} \frac{\partial w_i}{\partial n}(x,y) \, ds = 0 \quad i = 1, 2, ..., 5 \quad (9.114)
\]
where, as earlier in Section 9.2 on page 246, the number of elements identified with (attached to) each node is given by \(E_i\).

Let us now return to our example problem. We can write Eq. (9.114) in matrix form for the problem presented in Fig. 9.9 as
\[
\begin{bmatrix}
  a_{11} & a_{12} & a_{13} & a_{14} & 0 \\
  a_{21} & a_{22} & a_{23} & 0 & a_{25} \\
  a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\
  a_{41} & 0 & a_{43} & a_{44} & a_{45} \\
  0 & a_{52} & a_{53} & a_{54} & a_{55}
\end{bmatrix}
\begin{bmatrix}
  u_1 \\
  u_2 \\
  u_3 \\
  u_4 \\
  u_5
\end{bmatrix}
= \begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3 \\
  f_4 \\
  f_5
\end{bmatrix}. \quad (9.115)
\]
where the coefficients \(a_{ij}\) are given by
\[
a_{ij} = \sum_{e=1}^{E_i} \int_{\Omega_i} \left( -1.5 \frac{\partial \phi_j}{\partial x} \frac{\partial w_i}{\partial x} + 0.5 \frac{\partial \phi_j}{\partial y} \frac{\partial w_i}{\partial x} + 0.5 \frac{\partial \phi_j}{\partial x} \frac{\partial w_i}{\partial y} - 1.5 \frac{\partial \phi_j}{\partial y} \frac{\partial w_i}{\partial y} \right) \, dw. \quad (9.116)
\]
Note that each row of the matrix corresponds to a finite volume and each column contains information from all finite element nodes connected to that finite volume. Thus, for example, \(a_{15}\) is zero because the volume centered at node 1 does not have a connection to the finite element node 5.

At this point we realize, as in the first example, that we have only one unknown, namely node 3, because Dirichlet conditions are specified at all the other nodes and thus the values there are known. The resulting system of equations reduces to
\[
\begin{bmatrix}
  a_{31} & a_{32} & a_{33} & a_{34} & a_{35}
\end{bmatrix}
\begin{bmatrix}
  u_1 \\
  u_2 \\
  u_3 \\
  u_4 \\
  u_5
\end{bmatrix}
= \begin{bmatrix}
  f_3 \\
  f_4 \\
  f_5
\end{bmatrix}. \quad (9.117)
\]
or
\[
(a_{33}) \quad (u_3) = f_3 - (a_{31}) (1) - (a_{34}) (1). \quad (9.118)
\]
The coefficient \(a_{31}\) is of the form (where, as earlier, the first index, in this case 3, is the element number and the second index 1 is the connecting node).

\[
a_{31} = \sum_{e=1}^{E_3} \int_{\partial \Omega_3^3} \left( \left\{ 1.5 \frac{\partial \phi_1}{\partial x} - 0.5 \frac{\partial \phi_1}{\partial y} \right\} n_x + \left\{ 1.5 \frac{\partial \phi_1}{\partial x} - 0.5 \frac{\partial \phi_1}{\partial x} \right\} n_y \right) \, ds. \quad (9.119)
\]

### 9.3.3 Element Coefficient Matrices

Let us evaluate this coefficient for element 1. As in example problem 1, we need to consider two sides of the finite volume centered at node 3, namely the sides \(ab\) and \(bc\).
From
\[ \frac{\partial \phi_j}{\partial x} = \frac{y_{J+1} - y_{J+2}}{\det [P]} \]  
we obtain for \( \frac{\partial \phi_1}{\partial x} \)

\[ \frac{\partial \phi_1}{\partial x} = \frac{y_3 - y_4}{\det [P]} = \frac{0 - 1.4}{2} = \frac{-1}{\sqrt{2}} \]  

(9.121)

Similarly,

\[ \frac{\partial \phi_j}{\partial y} = \frac{x_{J+2} - y_{J+1}}{\det [P]} \]

which gives for \( \frac{\partial \phi_1}{\partial y} \)

\[ \frac{\partial \phi_1}{\partial y} = \frac{x_4 - x_3}{\det [P]} = \frac{1 - 1}{2} = 0 \]  

(9.123)

where \( \det [P] \) is given by Eq. (9.55).

Using Eqs. (9.22) and (9.23) on page 243 we obtain

\[ n_x|_{ab} = \frac{x_1 - x_3}{\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} = \frac{-\sqrt{2}}{\sqrt{2}} = -1 \]  

(9.124)

\[ n_x|_{bc} = \frac{x_4 - x_3}{\sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2}} = \frac{0}{\sqrt{2}} = 0 \]  

(9.125)

\[ n_y|_{ab} = \frac{y_1 - y_3}{\sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2}} = \frac{0}{\sqrt{2}} = 0 \]  

(9.126)

\[ n_y|_{bc} = \frac{y_4 - y_3}{\sqrt{(x_3 - x_4)^2 + (y_3 - y_4)^2}} = \frac{\sqrt{2}}{\sqrt{2}} = 1 \]  

(9.127)

We now substitute Eqs. (9.121)–(9.127) into Eq. (9.119) and obtain for the coef-
9.3. **FINITE VOLUME EXAMPLE PROBLEM TWO**

Coefficient $a_{31}$ of element 1

$$a_{31} = \int_{\partial \Omega_3} \left( \left\{ 1.5 \frac{\partial \psi}{\partial x} - 0.5 \frac{\partial \psi}{\partial y} \right\} n_x|_{ab} + \left\{ 1.5 \frac{\partial \phi}{\partial y} - 0.5 \frac{\partial \phi}{\partial x} \right\} n_y|_{ab} \right) ds$$

$$+ \int_{\partial \Omega_3} \left( \left\{ 1.5 \frac{\partial \psi}{\partial x} - 0.5 \frac{\partial \psi}{\partial y} \right\} n_x|_{bc} + \left\{ 1.5 \frac{\partial \phi}{\partial y} - 0.5 \frac{\partial \phi}{\partial x} \right\} n_y|_{bc} \right) ds$$

$$= \left[ \left\{ (1.5) \left( \frac{-1}{\sqrt{2}} \right) - (0.5) (0) \right\} (-1) \left( \frac{\sqrt{2}}{2} \right) + \left\{ (1.5) (0) - (0.5) \left( \frac{-1}{\sqrt{2}} \right) \right\} (0) \left( \frac{\sqrt{2}}{2} \right) \right]$$

$$+ \left\{ (1.5) \left( \frac{-1}{\sqrt{2}} \right) - (0.5) (0) \right\} (0) \left( \frac{\sqrt{2}}{2} \right) + \left\{ (1.5) (0) - (0.5) \left( \frac{-1}{\sqrt{2}} \right) \right\} (1) \left( \frac{\sqrt{2}}{2} \right)$$

$$= \frac{1.5}{2} + \frac{0.5}{2} = 1. \quad (9.128)$$

Notice that this coefficient, namely that for $a_{31}$ for element 1, is equivalent to the element coefficient at $I + 1, J$ in element 1. Returning to element coefficient matrix 1 in example problem 1, (the matrix noted as Eq. (9.74) on page 256) you will observe that the coefficient at $I + 1, J$ has the value 1.0. In fact all the coefficient values in all the element matrices are the same for the line-integral part of problem 1 and problem 2. In fact, since Table 9.1 on page 249 that relates the global coordinates to the local coordinates is the same for both example problems, the global matrix for the line-integral part is also the same for both problems. This is a result of the peculiar nature of the specific example problems considered here and would not generally be the case. As a result, it is not necessary to detail the coefficient evaluation of the matrix equation for this problem.

### 9.3.4 Evaluation of the Source Term

The remaining term has the following form

$$f_i = \int_{\Omega_i} Q w_i (x, y) d\Omega \quad (9.129)$$

which for constant $Q$ over each finite volume becomes

$$f_i = Q_i \int_{\Omega_i} w_i (x, y) d\Omega. \quad (9.130)$$

As an example, for element 1 and finite volume 3 we obtain

$$f_3 = Q_3 \int_{\Omega_3^e} w_3 (x, y) d\Omega$$

$$= Q_3 \det [P] \quad (9.131)$$

where $[P]$ is identified with element 1.

If we let $Q = 0$. we have for our equation

$$(a_{33})(u_3) = f_3 - a_{31} (1) - a_{34} (1)$$

$$(6)(u_3) = 0 - (1.5) (1) - (1.5) (1)$$

$$u_3 = 0.50 \quad (9.132)$$
which is the same answer we obtained in the original system before rotation. If we let $Q_3 = 1$ we obtain

\[
f_3 = \sum_{e=1}^{4} Q_3 \int_{\Omega_3} w_3(x, y) d\Omega
\]

\[
= 2 \quad (9.133)
\]

and

\[
(a_{33}) (u_3) = f_3 - a_{31} (1) - a_{34} (1)
\]

\[
-6u_3 = 2 - (1.5) (1) - (1) (1)
\]

\[
w_3 = 0.17. \quad (9.134)
\]

The above provides the protocol for solving time-independent problems using the finite-volume method for problems with variable coefficients and cross-derivatives. Extension to time dependent problems follows the strategy to be provided in Chapter 10.

### 9.4 Chapter Summary

The finite volume method is presented. A weighted residual strategy is used that is consistent with the finite element method. The formulation of the element coefficient matrices and their collection into the global coefficient matrix is analogous to the procedure used earlier with the finite element method. Two examples are presented in detail. The second is the equivalent of the first but for the rotation of the coordinate system so that a cross-derivative appears in the equation.

### 9.5 Problems

1. Fig. 9.10 illustrates five finite volumes and four finite elements. The finite volume associated with node 5 is square and shaded. The other finite volumes are triangular and centered at the nodes 1–4. Assume the element coefficient matrices are as follows ( $e$ is the number of the element)

\[
e = 1 \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (9.135)
\]

\[
e = 2 \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \quad (9.136)
\]

\[
e = 3 \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \quad (9.137)
\]

\[
e = 4 \begin{bmatrix} d_{11} & d_{12} & d_{13} \\ d_{21} & d_{22} & d_{23} \\ d_{31} & d_{32} & d_{33} \end{bmatrix}. \quad (9.138)
\]
The relationship between the global and local element node numbers is given below. The top row of numbers are the global node numbers. The second row of numbers contains the corresponding element node numbers for the first element, element 1 with element entries $a_{ij}$.

<table>
<thead>
<tr>
<th>Local</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>(a)</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>3</td>
<td></td>
<td>2</td>
<td>(b)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>2</td>
<td></td>
<td>1</td>
<td>(c)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>(d)</td>
</tr>
</tbody>
</table>

The global coefficient matrix will have 5 rows and 5 columns as shown below. Your task is to populate this matrix with the values from the four element matrices shown in Eqs. (9.135)–(9.138). One coefficient is completed as an example:

$$
\begin{bmatrix}
& & & & a_{13} + b_{12} \\
& & & & \\
& & & & \\
& & & & \\
& & & & \\
\end{bmatrix}
$$

Figure 9.10: The figure illustrates four triangular elements and five finite volumes. The finite volume associated with node 5 is shaded. The finite volume numbers correspond to the nodal numbers and the bold numbers are associated with the finite elements.

2. Please fill in the spaces in the columns with an X where appropriate. For example in the first row we ask what methods generate a $4 \times 4$ element coefficient matrix. Finite volume may do so and isoparametric always does, so a X is placed in the first and third columns. A blank is not counted one way or the other.
### Attribute

<table>
<thead>
<tr>
<th>Method Generation</th>
<th>FV</th>
<th>FE</th>
<th>Iso</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 × 4 element coefficient matrix</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Prism with polygonal cross section for weighting function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basis function as weighting function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basis functions defined on triangular subspaces</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basis functions defined on a natural coordinate system</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximation uses Green’s theorem (integration by parts)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elements are defined by straight lines connecting nodes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jacobian matrix in formulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conserves extensive property (for example mass)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type two (specified flux) boundary conditions are an integral part of the approximate equation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Employed the Dirac delta function in the equation formulation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximation of a derivative is constant over an element</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elements are in general, four-sided shapes with straight sides and no sides parallel</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Can easily provide small elements in the neighborhood of singularities, for example, a well</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creates a 4 × 4 element coefficient matrix</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uses method of weighted residuals in formulation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FV = finite volume, FE = finite element, Iso = isoparametric

3. Consider the finite volume mesh presented in Fig. 9.11 the following figure. The letters are the finite volumes, the italic numbers are the elements and the rest are nodes. Please answer the following:

(a) In the global matrix, how many non-zero coefficients will be found in the first line of the matrix?

(b) How many zero coefficients will be found in the matrix of the first line of the global matrix equation?

(c) What equations (row numbers) will contain information from node 1?

(d) Given the equations

\[
\sum_{e=1}^{E_i} \sum_{j=1}^{n} u_j \int_{\partial \Omega_e} \left( 2 \frac{\partial \phi_j(x,y)}{\partial x} n_x + \frac{\partial \phi_j(x,y)}{\partial y} n_y \right) ds \\
+ \sum_{e=1}^{E_i} \sum_{j=1}^{5} u_j \int_{\Omega_e} (\kappa \phi^e_j(x,y)) w_i(x,y) d\Omega - \int_{\Omega} Q w_i(x,y) d\Omega + \\
\int_{\partial \Omega} \left[ 2 \frac{\partial \hat{u}(x,y)}{\partial x} n_x + \frac{\partial \hat{u}(x,y)}{\partial y} n_y \right] w_i(x,y) ds \\
= 0 \quad i = 1, 2, \ldots, 5 
\]

(9.139)

what is the value of \( E_8 \)?
Figure 9.11: Definition sketch for question 9.5.


Chapter 10

Initial Boundary-Value Problems

In this section we will extend our two-space dimensional analysis to consider time-dependent problems. The equation we will consider is the convective-diffusion equations, that is

\[ a \frac{\partial u}{\partial t} + b \nabla^2 u + c \nabla u = 0 \]  

(10.1)

with boundary and initial conditions

\[ u (x, y, 0) = u_0 (x, y) \quad x, y \in \Omega \]  

(10.2)

\[ u (x, y, t) = u_b (x, y) \quad x, y \in \Omega D \]  

(10.3)

\[ \frac{\partial u (x, y, t)}{\partial n} = \frac{\partial u (x, y, t)}{\partial n_b} \quad x, y \in \Omega N. \]  

(10.4)

We now proceed in a fashion analogous to that presented in Section 6.5 on page 158. The first step is to define the approximation to \( u (x, y) \), namely \( \hat{u} (x, y) \)

\[ \hat{u} (x, y) = \sum_{j=1}^{N} u_j (t) \phi_j (x, y). \]  

(10.5)

Note that the unknown coefficients \( u_j (t) \) are now a function of time and the basis functions \( \phi_j (x, y) \) are a function only of space.

We now define a residual as

\[ a \frac{\partial \hat{u}}{\partial t} + b \nabla^2 \hat{u} + c \nabla \hat{u} = R (x, y, t). \]  

(10.6)

From Galerkin’s method we have

\[ \int_{\Omega} R (x, y, t) \phi_i (x, y) \, d\Omega = 0 \quad i = 1, \ldots, N. \]  

(10.7)

Substituting Eq. (10.6) into Eq. (10.7) we have

\[ \int_{\Omega} \left( a \frac{\partial \hat{u}}{\partial t} + b \nabla^2 \hat{u} + c \nabla \hat{u} \right) \phi_i (x, y) \, d\Omega = 0 \quad i = 1, \ldots, N. \]  

(10.8)

If we now substitute Eq. (10.5) into Eq. (10.8) we obtain

Copyright © 2018 John Wiley & Sons, Inc. Published by John Wiley & Sons, Inc.
\begin{align}
\sum_{j=1}^{N} \int_{\Omega} \left( a \frac{du_j}{dt} \phi_j(x, y) + bu_j \nabla^2 \phi_j(x, y) + cu_j \nabla \phi_j(x, y) \right) \phi_i(x, y) \, d\Omega &= 0 \quad i = 1, \ldots, N. \\
&\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad
\[
\int_{\Omega_e} \phi_j (x, y) \phi_i (x, y) \, d\Omega \\
= \int_{\Omega_e} \phi^1_i (x, y) \phi^1_{i+1} (x, y) \, d\Omega_e \\
= 2A_e \frac{1!1!0!}{(1 + 1 + 0 + 2)!} \\
= 2A_e \frac{1}{4!} \\
= \frac{A_e}{12}. \tag{10.19}
\]

For the case of \( i = j \) we have

\[
\int_{\Omega_e} \phi_j (x, y) \phi_j (x, y) \, d\Omega \\
= \int_{\Omega_e} \phi^1_i \phi^1_i \, d\Omega_e \\
= \int_{\Omega_e} \phi^2_i \, d\Omega_e \\
= 2A_e \frac{2!}{(2 + 0 + 0 + 2)!} \\
= 2A_e \frac{2!}{4!} \\
= \frac{A_e}{6}. \tag{10.20}
\]

The matrix equation (Eq. (10.12)) represents a set of \( N \) ordinary differential equations in time. We now approximate the time derivative using finite difference methods to obtain

\[
[M] \{u\}_{t+\Delta t} + [N] \frac{\{u\}_{t+\Delta t} - \{u\}_t}{\Delta t} + \{f\} = 0. \tag{10.21}
\]

Grouping terms we obtain

\[
\left( [M] + \frac{[N]}{\Delta t} \right) \{u\}_{t+\Delta t} = \frac{[N]}{\Delta t} \{u\}_t - \{f\} \tag{10.22}
\]

which can be simplified to read

\[
[C] \{u\}_{t+\Delta t} = \{g\} \tag{10.23}
\]

where

\[
[C] = \left( [M] + \frac{[N]}{\Delta t} \right) \tag{10.24}
\]

and

\[
\{g\} = \frac{[N]}{\Delta t} \{u\}_t - \{f\}. \tag{10.25}
\]
Thus we can solve Eq. (10.23) symbolically as

\[ \{u\}_{t+\Delta t} = [C]^{-1} \{g\}. \] (10.26)

### 10.1 Mass Lumping

While the form of the \([N]\) matrix as presented can be, and is used in practice, there are alternative strategies. One approach is to lump the matrix \([N]\). Lumping is defined as follows

\[ n_{ii} = \sum_{j=1}^{N} \int_{\Omega} a \phi_j (x, y) \phi_i (x, y) \, d\Omega \] (10.27)

\[ n_{ij} = 0 \quad i \neq j. \] (10.28)

Under this strategy the diagonal elements of \([N]\) are the only non-zero elements.

### 10.2 Chapter Summary

In this chapter we extended the finite difference and finite element methods introduced earlier to tackle the transient case.

### 10.3 Problems

1. Assume that you want to solve the equation, New

\[ a \frac{\partial u}{\partial t} + b \nabla^2 u = 0 \] (10.29)

using linear the approximation

\[ u (x, y, t) \approx \hat{u} (x, y, t) = \sum_{j=1}^{n} u_j \phi_j (x, y, t) \] (10.30)

(a) assuming you want to employ a linear approximation, that is one where there are no second degree terms in \(\phi_j (x, y, t)\), what two geometric forms can \(\phi_j (x, y, t)\) take?

(b) sketch one three-dimensional finite element of each form with views showing the \(x-y\) and \(x-t\) planes.

(c) explain why the bandwidth of the coefficient matrix increases when the formulation based upon \(\hat{u} (x, y, t) = \sum_{j=1}^{n} u_j \phi_j (x, y, t)\) is used rather than \(\hat{u} (x, y, t) = \sum_{j=1}^{n} u_j (t) \phi_j (x, y)\).
Bibliography


Chapter 11

Boundary-Value Problems in Three Space

In this chapter we will extend the concepts introduced earlier for the solution of problems in two space dimensions to problems in three space dimensions. We begin with the finite difference approximations.

11.1 Finite Difference Approximations

The point of departure for the approximation of three-dimensional problems using finite difference methods is the material found in Section 7 on page 169. Eq. (7.5) which represented a function in the \((x, y)\) coordinate system using Lagrange polynomials is now expanded to consider three space dimensions and is written as

\[
u(x, y, z) = \sum_{i=0}^{n} \sum_{j=0}^{m} \sum_{k=0}^{n} \ell_i^n(x) \ell_j^m(y) \ell_k^p(z) u_{i,j,k} + E_{nmp}\quad (11.1)
\]

where \(k\) is the index and \(p\) is the order of the Lagrange polynomial in the \(z\) direction.

We can now obtain an approximation of the derivatives we normally encounter in the solution of the spatially-dependent partial-differential equations of interest to us. For the first order derivatives we have

\[
\left. \frac{\partial u}{\partial x} \right|_{i,j,k} = \frac{1}{2\Delta x} [u_{i+1,j,k} - u_{i-1,j,k}] + \frac{\partial}{\partial x} E_{i,j,k} \quad (11.2)
\]

\[
\left. \frac{\partial u}{\partial y} \right|_{i,j,k} = \frac{1}{2\Delta y} [u_{i,j+1,k} - u_{i,j-1,k}] + \frac{\partial}{\partial y} E_{i,j,k} \quad (11.3)
\]

\[
\left. \frac{\partial u}{\partial z} \right|_{i,j,k} = \frac{1}{2\Delta z} [u_{i,j,k+1} - u_{i,j,k-1}] + \frac{\partial}{\partial z} E_{i,j,k} \quad (11.4)
\]

and for the second derivatives we obtain

\[
\left. \frac{\partial^2 u}{\partial x^2} \right|_{i,j,k} = \frac{1}{\Delta x^2} [u_{i-1,j,k} - 2u_{i,j,k} + u_{i+1,j,k}] + \frac{\partial^2}{\partial x^2} E_{i,j,k} \quad (11.5)
\]

\[
\left. \frac{\partial^2 u}{\partial y^2} \right|_{i,j,k} = \frac{1}{\Delta y^2} [u_{i,j-1,k} - 2u_{i,j,k} + u_{i,j+1,k}] + \frac{\partial^2}{\partial y^2} E_{i,j,k} \quad (11.6)
\]
\[
\frac{\partial^2 u}{\partial z^2}|_{i,j,k} = \frac{1}{\Delta z^2} \left[ u_{i,k-1} - 2u_{i,j,k} + u_{i,j+1,k+1} \right] + \frac{\partial^2}{\partial z^2} E_{i,j,k}. \tag{11.7}
\]

Now, let us use these approximations in the equation
\[
\begin{align*}
\alpha_{xx} (x, y, z) \frac{\partial^2 u (x, y, z)}{\partial x^2} + \alpha_{yy} (x, y, z) \frac{\partial^2 u (x, y, z)}{\partial y^2} + \alpha_{zz} (x, y, z) \frac{\partial^2 u (x, y, z)}{\partial z^2} \\
- \beta_x (x, y, z) \frac{\partial u (x, y, z)}{\partial x} - \beta_y (x, y, z) \frac{\partial u (x, y, z)}{\partial y} - \beta_z (x, y, z) \frac{\partial u (x, y, z)}{\partial z} + u (x, y, z) &= 0. \tag{11.8}
\end{align*}
\]

where \(\alpha_{xx}, \alpha_{yy}, \alpha_{zz}, \beta_x, \beta_y, \) and \(\beta_z\) are known coefficients which may be a function of space. Substitution of the approximations for the node \((i\Delta x, j\Delta y, k\Delta z)\) in Eq. (11.8) yields
\[
\begin{align*}
&\frac{\alpha_{xx}}{\Delta x^2} \left[ u_{i-1,j,k} - 2u_{i,j,k} + u_{i+1,j,k} \right] \\
+ &\frac{\alpha_{yy}}{\Delta y^2} \left[ u_{i,j-1,k} - 2u_{i,j,k} + u_{i,j+1,k} \right] \\
+ &\frac{\alpha_{zz}}{\Delta z^2} \left[ u_{i,j,k-1} - 2u_{i,j,k} + u_{i,j,k+1} \right] \\
- &\frac{\beta_x}{2\Delta x} \left[ u_{i+1,j,k} - u_{i-1,j,k} \right] \\
- &\frac{\beta_y}{2\Delta y} \left[ u_{i,j+1,k} - u_{i,j-1,k} \right] \\
- &\frac{\beta_z}{2\Delta z} \left[ u_{i,j,k+1} - u_{i,j,k-1} \right] + u_{i,j,k} \\
= &0. \tag{11.9}
\end{align*}
\]

The computational molecule for this finite difference approximation is seen in Fig. 11.1. The reference point is at \((i\Delta x, j\Delta y, k\Delta z)\). It is a straightforward extension of the two-dimensional finite difference in space molecule; the only change is the number of subscripts and the addition of the vertical space dimension. In contrast to the two-dimensional case, each equation (row) of the matrix equation will now have seven non-zero coefficients organized into seven diagonal bands. The matrix is generated by using the template of Eq. (11.9) for each node in the finite difference mesh. Boundary conditions are handled in the same manner as for the two-dimensional problem.

### 11.2 Finite Element Approximations

The extension of the two-dimensional finite element approximation method can be approached in two ways, the choice depending upon the nature of physical system being considered. We will illustrate the methodology using the example equation presented above, that is
\[
\begin{align*}
\alpha_{xx} (x, y, z) \frac{\partial^2 u (x, y, z)}{\partial x^2} + \alpha_{yy} (x, y, z) \frac{\partial^2 u (x, y, z)}{\partial y^2} + \alpha_{zz} (x, y, z) \frac{\partial^2 u (x, y, z)}{\partial z^2} \\
- \beta_x (x, y, z) \frac{\partial u (x, y, z)}{\partial x} - \beta_y (x, y, z) \frac{\partial u (x, y, z)}{\partial y} - \beta_z (x, y, z) \frac{\partial u (x, y, z)}{\partial z} + u (x, y, z) &= 0. \tag{11.10}
\end{align*}
\]
11.2. FINITE ELEMENT APPROXIMATIONS

The first step is to employ the Galerkin method of approximation to obtain the weighted residual formulation. To do this we first define a residual by substituting the sought-after function \( u(x, y, z) \) with an approximate value \( \hat{u}(x, y, z) \). The result is

\[
\alpha_{xx}(x, y, z) \frac{\partial^2 \hat{u}(x, y, z)}{\partial x^2} + \alpha_{yy}(x, y, z) \frac{\partial^2 \hat{u}(x, y, z)}{\partial y^2} + \alpha_{zz}(x, y, z) \frac{\partial^2 \hat{u}(x, y, z)}{\partial z^2}
\]

\[-\beta_x(x, y, z) \frac{\partial \hat{u}(x, y, z)}{\partial x} - \beta_y(x, y, z) \frac{\partial \hat{u}(x, y, z)}{\partial y} - \beta_z(x, y, z) \frac{\partial \hat{u}(x, y, z)}{\partial z} + \hat{u}(x, y, z) \equiv R(x, y, z). \tag{11.11}
\]

The second step is to weight the residual defined in Eq. (11.11) by a test function \( \phi_i(x, y, z) \), the nature of which will be explained shortly, and integrate over the domain of interest. The region \( \Omega(x, y, z) \) is now three-dimensional (see Fig. 11.2), rather than the two-dimensional form shown in Fig. 8.9 on page 8.9. This process yields

\[
\int_{\Omega} R(x, y, z) \phi_i(x, y, z) = 0 \quad i = 1, \ldots, N. \tag{11.12}
\]

The approximating function \( \hat{u}(x, y, z) \) is now approximated as

\[
\hat{u}(x, y, z) = \sum_{j=1}^{N} u_j \phi_j(x, y, z) \tag{11.13}
\]

where the functions \( \phi_j(x, y, z) \) are the same as those used as the weighting functions in Eq. (11.12). Substitution of Eqs. (11.13) and (11.11) into Eq. (11.12) yields
Figure 11.2: Three-dimensional region as encountered in three-dimensional finite element analysis.

\[ \sum_{j=1}^{N} u_j \int_{\Omega} \left[ \alpha_{xx}(x, y, z) \frac{\partial^2}{\partial x^2} \phi_j(x, y, z) + \alpha_{yy}(x, y, z) \frac{\partial^2}{\partial y^2} \phi_j(x, y, z) + \alpha_{zz}(x, y, z) \frac{\partial^2}{\partial z^2} \phi_j(x, y, z) \right. \\
+ \left. \alpha_{xy}(x, y, z) \frac{\partial}{\partial x} \frac{\partial}{\partial y} \phi_j(x, y, z) + \alpha_{yz}(x, y, z) \frac{\partial}{\partial y} \frac{\partial}{\partial z} \phi_j(x, y, z) + \alpha_{xz}(x, y, z) \frac{\partial}{\partial x} \frac{\partial}{\partial z} \phi_j(x, y, z) \right] \right. \\
- \left. \beta_x(x, y, z) \frac{\partial}{\partial x} \phi_j(x, y, z) - \beta_y(x, y, z) \frac{\partial}{\partial y} \phi_j(x, y, z) - \beta_z(x, y, z) \frac{\partial}{\partial z} \phi_j(x, y, z) \right] d\Omega \\
= 0 \quad i = 1, 2, ..., N. \quad (11.14) \]

The next step is to introduce the finite element notation which replaces the domain \( \Omega \) with a series of \( E \) finite elements in Eq. (11.14). Application of Green’s theorem changes the form of the second-order terms in this equation. In the process, a line integral along the boundary \( \partial \Omega \) is created. The result is
11.2. FINITE ELEMENT APPROXIMATIONS

\[ \sum_{j=1}^{N} u_j \int_{\Omega} \left[ \alpha_{xx} (x, y, z) \frac{\partial}{\partial x} \phi_j (x, y, z) \frac{\partial}{\partial x} \phi_i (x, y, z) \\
+ \alpha_{yy} (x, y, z) \frac{\partial}{\partial y} \phi_j (x, y, z) \frac{\partial}{\partial y} \phi_i (x, y, z) \\
+ \alpha_{zz} (x, y, z) \frac{\partial}{\partial z} \phi_j (x, y, z) \frac{\partial}{\partial z} \phi_i (x, y, z) \\
+ \beta_x (x, y, z) \frac{\partial}{\partial x} \phi_j (x, y, z) + \beta_y (x, y, z) \frac{\partial}{\partial y} \phi_j (x, y, z) \\
+ \beta_z (x, y, z) \frac{\partial}{\partial z} \phi_j (x, y, z) \right] \ d\Omega_e - \\
\int_{\partial \Omega_e} a_n \frac{\partial \hat{u}(x,y,z)}{\partial n} \phi_i (x, y, z) \ d\omega = 0 \quad i = 1, 2, ..., N \]  

(11.15)

We now introduce the concept of integration over finite elements where \( e \) denotes the element number. This yields

\[ \sum_{e=1}^{E} \sum_{j=1}^{N} u_j \int_{\Omega_e} \left[ \alpha_{xx_e} (x, y, z) \frac{\partial}{\partial x} \phi_j (x, y, z) \frac{\partial}{\partial x} \phi_i (x, y, z) \\
+ \alpha_{yy_e} (x, y, z) \frac{\partial}{\partial y} \phi_j (x, y, z) \frac{\partial}{\partial y} \phi_i (x, y, z) \\
+ \alpha_{zz_e} (x, y, z) \frac{\partial}{\partial z} \phi_j (x, y, z) \frac{\partial}{\partial z} \phi_i (x, y, z) \\
+ \beta_x (x, y, z) \frac{\partial}{\partial x} \phi_j (x, y, z) + \beta_y (x, y, z) \frac{\partial}{\partial y} \phi_j (x, y, z) \\
+ \beta_z (x, y, z) \frac{\partial}{\partial z} \phi_j (x, y, z) \right] \ d\Omega_e - \\
\int_{\partial \Omega_e} a_n \frac{\partial \hat{u}(x,y,z)}{\partial n} \phi_i (x, y, z) \ d\omega = 0 \quad i = 1, 2, ..., N \]  

(11.16)

where the subscript \( e \) on the coefficients \( \alpha \) indicate that they are considered as element-wise constant values. Imbedded in Eq. (11.16) are the \( \phi_i (x, y, z) \) and \( \phi_i (x, y, z) \) basis and weighting functions respectively.

The remaining task is to define the functions \( \phi_i (x, y, z) \) and \( \phi_j (x, y, z) \) which are defined element-wise as was done in Section 8.2.1 on page 196. There natural choice for three dimensional simulation is the three-dimensional extension of the triangle, the tetrahedron. A typical tetrahedral element is seen in Fig. 11.3.

To formulate the appropriate tetrahedral basis functions we follow the same procedure we did for triangles (see Section 8.2.1 ). We begin by expressing the functions \( \phi_i (x, y, z) \) using the expression definition of athe triangular-based pyramid, that is,

\[ \phi_i (x, y, z) = a_i + b_i x + c_i y + d_i z \]

and impose the requirements of a basis function, namely that \( \phi_i (x, y, z) \) must be unity at the node for which it is defined (here node \( i \)) and zero at all other nodes. Thus we can write the following set of equations for \( \phi_i (x, y, z) \), where \( I \) is the local element index as seen in Fig. 11.3
The solution to Eq. (11.17) is given as

\[
\phi_I(x, y, z) = \frac{1}{6V} (a_I + b_I x + c_I y + d_I z)
\]  

(11.18)

where

\[
a_I = \det \begin{bmatrix} x_{I+1} & y_{I+1} & z_{I=1} \\ x_{I+2} & y_{I+2} & z_{I=2} \\ x_{I+3} & y_{I+3} & z_{I=3} \end{bmatrix}
\]

(11.19)

\[
b_I = \det \begin{bmatrix} 1 & y_{I+1} & z_{I=1} \\ 1 & y_{I+2} & z_{I=2} \\ 1 & y_{I+3} & z_{I=3} \end{bmatrix}
\]

(11.20)

\[
c_I = \det \begin{bmatrix} x_{I+1} & 1 & z_{I=1} \\ x_{I+2} & 1 & z_{I=2} \\ x_{I+3} & 1 & z_{I=3} \end{bmatrix}
\]

(11.21)

\[
d_I = \det \begin{bmatrix} x_{I+1} & y_{I+1} & 1 \\ x_{I+2} & y_{I+2} & 1 \\ x_{I+3} & y_{I+3} & 1 \end{bmatrix}
\]

(11.22)

and

\[
6V = \det \begin{bmatrix} 1 & x_I & y_I & z_I \\ 1 & x_{I+1} & y_{I+1} & z_{I=1} \\ 1 & x_{I+2} & y_{I+2} & z_{I=2} \\ 1 & x_{I+3} & y_{I+3} & z_{I=3} \end{bmatrix}
\]

(11.23)

The equations for \(\phi_{I+2}, \phi_{I+3}\), and \(\phi_{I+3}\) can be obtained by rotating through the remaining nodes and populating the various determinants using the right-hand rule. As an example,

\[
a_{I+1} = \begin{bmatrix} x_{I+2} & y_{I+2} & z_{I+2} \\ x_{I+3} & y_{I+3} & z_{I+3} \\ x_I & y_I & z_I \end{bmatrix}
\]

(11.24)

Recall that the area of an element is related to its determinant. (see p. 199)

Remember also that with triangles we had to number the nodes (corners) in a counterclockwise manner to assure that the determinant was positive. Similarly in the case of tetrahedral elements the numbering matters. One scheme that assures that the determinant in Eq. (11.23) will be positive is the following: take any node location and call it \(I\). Now select a side and number the nodes increasing counterclockwise. For example in Fig. 11.3 we number the side as \(I, I+1, I+2\), and the fourth node, off the plane of the others, on the opposite side of the tetrahedral element, is \(I+3\) In the evaluation process, the values of \(I, I+1, I+2\), and \(I+3\) would be replaced by global nodal values with their associated locations when populating the determinants noted above.

The rest of the process of formulating and solving the discrete equations follows the process outlined in Section 8.2.1. The extension to the transient case is also achieved
using the same procedure as outlined for the triangular element case in Section 10 on page 273.

There remains to be considered the term

$$\int_{\partial \Omega_e} a_{n_e} \frac{\partial \hat{\nu}}{\partial n} \bigg|_e \phi_i(x, y, z) \, d\omega.$$  \hspace{1cm} (11.25)

In the three dimensional formulation this is an integral over the boundary of the area of interest, considered on element at a time. Consequently this integration is the same as that considered for triangular elements. In this case, the $\phi_i(x, y, z)$ term is a triangle on the surface of the domain. The integrated values are then allocated to each of the three nodes per element.

Integrations in three dimensions can be achieved using a formula analogous to that used for triangles. The formula is

$$\int_{\Omega_e} \phi_I^{m_I} \phi_{I+1}^{m_{I+1}} \phi_{I+2}^{m_{I+2}} \phi_{I+3}^{m_{I+3}} \, d\Omega_e = 6V \frac{m_I!m_{I+1}!m_{I+2}!m_{I+3}!}{m_I + m_{I+1} + m_{I+2} + m_{I+3} + 3}$$ \hspace{1cm} (11.26)

where, as earlier, $m_I$ is the power to which the polynomial is raised.

### 11.3 Chapter Summary

In this chapter we extended our earlier analysis of finite difference and finite element from two space dimensions to three. The fundamental formulation in three dimensions for finite difference is very similar to that for two dimensions. In the case of finite elements, it is necessary to introduce a new finite element based upon the tetrahedron. Again, the basic theory is analogous to that developed for two dimensions. Although we did not consider collocation or finite volume extensions to three dimensions, such extensions introduce no new concepts.
Bibliography

http://www.colorado.edu/engineering/cas/courses.d/AFEM.d/Home.html.


Chapter 12

Nomenclature

Latin alphabet

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\overline{2G}$</td>
<td>line segment connection node 2 to point $G$</td>
</tr>
<tr>
<td>$a_i$</td>
<td>known constant</td>
</tr>
<tr>
<td>$a_{ij}$</td>
<td>matrix coefficient of matrix $[A]$</td>
</tr>
<tr>
<td>$a_{xx}, a_{xy}, a_{yx}, a_{yy}$</td>
<td>components of coefficient tensor</td>
</tr>
<tr>
<td>$\overline{AF}$</td>
<td>line segment connecting point $A$ to point $F$</td>
</tr>
<tr>
<td>$b$</td>
<td>constant</td>
</tr>
<tr>
<td>$b_j$</td>
<td>known constant</td>
</tr>
<tr>
<td>$b_{ij}$</td>
<td>coefficient in matrix $[B]$</td>
</tr>
<tr>
<td>$c$</td>
<td>constant</td>
</tr>
<tr>
<td>$c(x,t)$</td>
<td>concentration</td>
</tr>
<tr>
<td>$d$</td>
<td>a constant</td>
</tr>
<tr>
<td>$e_i$</td>
<td>value in vector $[e]$</td>
</tr>
<tr>
<td>$e$</td>
<td>element number index when used as superscript</td>
</tr>
<tr>
<td>$D$</td>
<td>dispersion coefficient</td>
</tr>
<tr>
<td>$f(x)$</td>
<td>function of coordinate $x$</td>
</tr>
<tr>
<td>$\tilde{f}(x)$</td>
<td>approximation of function $f(x)$</td>
</tr>
<tr>
<td>$f(x_i)$</td>
<td>function $f(x)$ evaluated at location $x_i$</td>
</tr>
<tr>
<td>$\overline{EA}$</td>
<td>line segment connecting point $E$ to point $A$</td>
</tr>
<tr>
<td>$\overline{FB}$</td>
<td>line segment connecting point $F$ to point $B$</td>
</tr>
<tr>
<td>$g(x)$</td>
<td>function of coordinate $x$</td>
</tr>
<tr>
<td>$\overline{GB}$</td>
<td>line segment connecting point $G$ to point $B$</td>
</tr>
<tr>
<td>$\overline{GC}$</td>
<td>line segment connecting point $G$ to point $C$</td>
</tr>
<tr>
<td>$h_i^0$</td>
<td>Hermite cubic basis function of type 0 at node $i$</td>
</tr>
<tr>
<td>$h_i^1$</td>
<td>Hermite cubic basis function of type 1 at node $i$</td>
</tr>
<tr>
<td>$i$</td>
<td>unit vector in $x$ coordinate direction</td>
</tr>
<tr>
<td>$j$</td>
<td>unit vector in $y$ coordinate direction</td>
</tr>
<tr>
<td>$P_j^n$</td>
<td>Lagrange polynomial of degree $n$ defined at node $j$</td>
</tr>
<tr>
<td>$L$</td>
<td>characteristic length and also natural coordinate system coordinate</td>
</tr>
<tr>
<td>$m_I$</td>
<td>power to which the function $\phi_{I_i}$ is raised e.g. $\phi_{I_i}^{m_I}$</td>
</tr>
<tr>
<td>$m$</td>
<td>degree of Lagrange polynomial in time</td>
</tr>
<tr>
<td>$n$</td>
<td>normal coordinate in $(n, \tau)$ coordinate system</td>
</tr>
<tr>
<td>$O(x)$</td>
<td>order of $x$</td>
</tr>
<tr>
<td>$P_e$</td>
<td>Peclet number</td>
</tr>
<tr>
<td>$P_n(x)$</td>
<td>polynomial of degree $n$ defined in the $x$ coordinate direction</td>
</tr>
</tbody>
</table>
CHAPTER 12. NOMENCLATURE

$Q$  
flow rate entering a pipe

$u(x)$  
a function of coordinate $x$

$\hat{u}(x)$  
approximation of $u(x)$

$u_i$  
value of variable $u$ at node location $i$

$v$  
velocity

$u_{i,j}$  
value of $u$ at location $x+i\Delta x, j+j\Delta y$

$n$  
degree of Lagrange polynomial in space

$x_I$  
coordinate location for local element index $I$

$y_I$  
coordinate location for local element index $I$

$k_i$  
values associated with Runge–Kutta formulations

$T$  
temperature and also indicates transpose of a matrix when used as a superscript

$w_i(x)$  
weighting function defined in $x$ coordinate system

$w(x,y)$  
weighting function defined in $x-y$ coordinate system

$R(x)$  
residual defined in $x$ coordinate direction

$R(x,y)$  
residual defined in $x$ coordinate direction

$x_i$  
nodal location $i$ in along the $x - y$ coordinate system

Greek alphabet

$\beta$  
real coefficient

$\beta_i$  
constant

$\delta_{ij}$  
Kronecker delta

$\Gamma$  
diffusivity

$\delta(x - x_i)$  
Dirac delta function defined for location $\bar{x}_i$

$\varepsilon$  
measure of error

$\xi$  
natural coordinate

$\eta$  
natural coordinate

$\theta$  
trigonometric angle

$\kappa$  
coefficient

$\lambda_i$  
eigenvalue

$\Pi_j^i$  
multiplication operator

$\rho$  
density

$\sigma$  
coordinate direction for integration

$\tau$  
along a two-dimensional element boundary

unit vector in $\tau$ coordinate direction

$\tau$  
tangential coordinate in $(n, \tau)$ coordinate system

$\phi_j(x,y)$  
triangular basis function defined at node $j$

$\chi$  
local coordinate

$\psi_j(\xi, \eta)_i$  
basis function defined at node $j$

Matrices

$[A]$  
space matrix

$\{e\}$  
error vector

$[J]$  
Jacobian matrix

$[P]$  
matrix of nodal location information

used in defining element matrix coefficients

Expressions

$\bar{e}_i \equiv u_i - \bar{u}_i$  
error between function $u$ and its approximation $\bar{u}$

$E(\theta) \equiv f(\theta) - P_2(\theta)$  
error as a function of coordinate $\theta$

$h \equiv x_i - x_{i-1}$  
incremental distance
\[ L(\cdot) \equiv \frac{d^2(\cdot)}{dx^2} + \frac{d(\cdot)}{dx} + 0.5 (\cdot) \quad \text{mathematical operator as defined} \]

\[ p_n(x) = \Pi_{i=0}^{i=n}(x - x_i) \]

\[ \Delta x \equiv x_{i+1} - x_i \quad \text{spatial increment in coordinate } x \]

\[ \Delta y \equiv y_{j+1} - y_j \quad \text{spatial increment in coordinate } y \]

\[ \Delta t \equiv t - t_{i-1} \quad \text{increment in time } t \]

\[ \partial \Omega \quad \text{line or surface of } \Omega \text{ when } \Omega \text{ is an area or volume respectively} \]

\[ \partial \Omega_i^c \quad \text{line emanating from node } I \text{ in element } e \]

\[ \partial \Omega_D \quad \text{segment of boundary of } \Omega \text{ designated as a Dirichlet boundary condition} \]

\[ \partial \Omega_N \quad \text{segment of boundary of } \Omega \text{ designated as a Neumann boundary condition} \]
Index

numerical integration
  common formulas, 56

approximation of a function, 4

backward difference approximation, 36, 161
backward in time approximation, 147
basis functions, 85, 95, 184, 217, 273
  Lagrange, 85
  triangular, 196
boundary conditions, 107, 120, 144, 163, 273
  Dirichlet, 149
  Neumann, 206
boundary-value problems in three space coordinates, 279
centered in space time approximation, 147
chapter summary
  chapter eight, 230
  chapter eleven, 285
  chapter five, 133
  chapter four, 78
  chapter nine, 266
  chapter one, 24
  chapter seven, 175
  chapter six, 162
  chapter ten, 276
  chapter three, 62
  chapter two, 46
collocation method, 123, 124
  first-order equations, 123
  second-order equations, 126
collocation points, 124
computational molecule, 143
conditional stability, 162
consistency
  definition, 72
convergence
  definition, 69
coordinate transformation function, 216, 219
coordinates
  global, 196
  local, 211
  natural, 211
cross derivative, 262
  finite volume formulation, 241
difference approximation
  first derivative, 102
  second derivative, 279
difference equation, 73
Dirac delta function, 113, 124, 243
Dirichlet boundary condition, 150, 263
discretization, 2
eigenvalue, 151
eigenvector, 151
element coefficient matrix, 92, 105
  for triangular elements, 202
  linear elements, 120
eroof of the approximation, 4
  finite difference formulae, 36
  first-order derivative, 42, 43
  Hermite, 23
  quadratic Lagrange polynomial, 9
Euler forward integration formula, 68
Euler forward integration method example, 68
finite difference approximation, 35, 169
  in three space variables, 279
  two space dimensions, 141
finite difference interpretation
  of the first-order Galerkin approximation, 102
  of the second-order Galerkin approximation, 111
finite element
  approximation in three space variables, 280
  approximation over triangles, 195
  example problem, 199

Copyright © 2018 John Wiley & Sons, Inc. Published by John Wiley & Sons, Inc.

293
approximations over rectangles, 181
Galerkin approximation in time, 158, 161
forward difference approximation, 160
isoparametric, 211
basis functions, 217
tetrahedral, 283
triangular, 196
finite volume
application of Green’s theorem, 241
approximation, 239
approximation to second-order equations, 112
example problem, 262
example problem 1, 246
element coefficient matrices, 248
evaluation of the area integral, 256
evaluation of the line integral, 249
global matrix assembly, 260
weighted residual formulation, 246
example problem 2
element coefficient matrices, 263
evaluation of the source term, 265
problem definition, 262
weighted residual formulation, 262
finite difference interpretation, 122
finite difference interpretation for second-order equations, 122
finite element and polygon domain, 240
interpretation in terms of finite difference method, 93
mesh, 240
method, 133
normal-tangential boundary condition, 242
vector normalization, 243
weighted residual formulation, 240
weighting function, 84
weighting function in two space dimensions, 240
first derivative difference approximation, 41, 102
flux type boundary condition, 106
forward difference, 35, 160
Galerkin method, 94, 159, 200
finite element approximation in time, 158
for first-order equations, 94
for second-order equations, 102
time approximation
backward difference approximation, 161
Gauss integration, 61
Gauss points, 124
Gauss quadrature, 59, 124
global coefficient matrix, 93, 120, 204, 265
global coordinates, 196
global system, 89
Green’s theorem, 184, 201, 241, 274
Hermite polynomials, 20, 127
development of expression for, 20
implicit approximation, 148
incremental distance $h$, 35
initial boundary value problems, 273
mass lumping, 276
initial conditions, 67, 144, 163, 273
initial value problems, 67
inner product, 84
integration
finite volume method, 251
for isoparametric finite elements, 227
formula for triangular element, 198
three dimensional finite elements (tetrahedral elements), 285
integration by parts, 103, 117, 159, 184
interpolant, 7
interpolation, 1
interpretation of finite volume method in terms of the finite difference method, 93
isoparametric finite element, 211
basis functions, 217, 218
coordinate transformation functions, 216
example problem, 223
Gauss-quadrature integration, 227
Jacobian calculation, 219
matrix assembly, 228
natural coordinate system, 211
relationship between natural and global coordinates, 225
two-dimensional natural coordinate system, 212
Jacobian, 219
calculation of, 219
Jacobian matrix, 221
Kronecker delta, 5

Lagrange basis functions, 86
Lagrange interpolation, 4, 8, 55
Lagrange polynomials, 35, 55, 96, 141, 162, 170
  linear, 5, 85
  local coordinate system, 107
  quadratic, 8
linear Lagrange polynomial, 207
local coefficient matrix, 187
local coordinates, 88, 98, 104, 191
local Lagrange polynomials, 120
local system, 90
local — global coefficient transformation matrix, 189
locally defined Lagrange polynomials, 96
mass lumping, 276
matrix assembly isoparametric elements, 228
matrix method of stability analysis, 149
method of weighted residuals, 83, 112, 123, 159, 183, 240
multiple elements, 14
natural coordinate systems
  one dimensional natural coordinate systems, 211
  two dimensional natural coordinate systems, 212
  use in isoparametric integration, 222
Neumann boundary condition, 106, 206
Newton—Cotes quadrature formulae, 55
node, 4
numerical differentiation, 33
general theory, 33
numerical integration, 55
numerical solution of boundary value problems
  in two space dimensions
    finite difference approximations, 169
    finite element approximations, 180
    finite-difference approximations, 279
three-point difference formulae, 40
trapezoidal rule, 56
triangular basis functions
  formulation, 196
tetrahedral element, 284
triangular element, 284
two dimensional natural coordinate systems, 212
two-dimensional polynomial approximation, 139, 169
two-point difference formulae, 34
two-point formulae from Taylor Series, 37
weighting function, 83, 106, 124, 184
  finite volume method, 84
  Galerkin’s method, 95
Weirstraus Approximation Theorem, 3

numerical solution of ordinary differential equations, 67

II, definition of, 6
quadratic Lagrange polynomial, 8
residual error, 83
round-off error, 150
Runge–Kutta
g-geometric interpretation, 77
second-order form, 75
Runge–Kutta
type formulae, 75
second type or Neumann boundary value problem, 206
second-order derivatives, 43
σ coordinate system, 207
Simpson’s rule, 57
stability, 73, 151
  example of, 153
  example simulation, 156
stability of finite difference approximations
to boundary value problems, 149
steady state behavior, 169
subdomain method, 84
tetrahedral element, 284
three space dimensions, 279
  finite element approximations, 280
  finite-difference approximations, 279
three-point difference formulae, 40
trapezoidal rule, 56
triangular basis functions
  formulation, 196
tetrahedral element, 284
triangular element, 284
two dimensional natural coordinate systems, 212
two-dimensional polynomial approximation, 139, 169
two-point difference formulae, 34
two-point formulae from Taylor Series, 37
weighting function, 83, 106, 124, 184
  finite volume method, 84
  Galerkin’s method, 95
Weirstraus Approximation Theorem, 3