Hebron University
Faculty of Graduate Studies
Mathematics Department

# Approximating Linear Poisson's Equation Solutions Using the Finite Element Method-Analysis and Computation

Prepared by

Shahd Khalil Skafi

Supervisor

Dr. Hasan Almanasreh

This Thesis is Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Mathematics, Faculty of Graduate Studies, Hebron University, Hebron, Palestine.

2023

**Approximating Linear Poisson's Equation Solutions Using the Finite Element Method-Analysis and Computation**

**By**
**Shahd Khalil Skafi**

**This thesis was defended successfully on 27/04/2023 and approved by:**

| **Committee Members:** | | **Signature** |
|---|---|---|
| • Dr. Hasan Almanasreh | Supervisor | ................. |
| • Dr. Mahmoud Shalalfeh | Internal Examiner | ................. |
| • Dr. Naeem Alkoumi | External Examiner | ................. |

# Declaration

The work provided in this thesis, unless otherwise referenced, is the result of the researcher's work, and has not been submitted elsewhere for any other degree or qualification.

Shahd Skafi

Signature:________________ Date:_____________

# Dedication

*To Tareq, Khalil, Nahla, and Maria*

# Acknowledgements

I am grateful to my supervisor Dr. Hasan Almanasreh, for his continuous support and encouragement, which enabled me to write and understand the subject.
I acknowledge the support of Hebron University for this work, and I appreciate the effort of all respected staff in the Department of Mathematics and thank them for it.
I extend my sincere thanks to the distinguished members of the discussion committee for the valuable proposals they will give to this thesis, with a view to correcting and advancing it.

# Abstract

This thesis treats Poisson's equation and its approximate solution using the finite element method. For this purpose, a general theory of the finite element method and its strategy are explained briefly. Then, Poisson's equation, its derivation and applications are explained in details. The finite element method is then applied to approximate the solution of Poisson's equation in one and two dimensional spaces. In this essence, error analysis is studied for the Poisson's equation in both categories, a posteriori and a priori error estimates.

In addition, the thesis discusses the numerical solution for the Poisson's equation throughout examples in one and two dimensions using the MATLAB software.

…

# Contents

*Chapter 1*

# Differential equations and the FEM

The Finite Element Method, FEM, is a numerical technique used to approximate solutions of differential equations. It was originated from the need of solving complex elasticity and structural analysis problems in Civil and Mechanical engineering and mathematical physics [**1, 22**]. Typical problem areas of interest include structural analysis heat transfer, fluid flow, mass transports and electromagnetic potential. The formulation of a problem using FEM produce a system of algebraic equation, in which the unknown function over the domain is divided into smaller parts called finite elements. The equations that model these finite elements are then assembled into a large system of equations that models the problem over the entire domain. Based on calculus of variation, the FEM uses variational method to approximate a solution by minimizing an associated error function [**18, 7, 21**].

## 1.1 History of the analysis of the finite element

Ritz [**1, 9, 11**] was the first who proposed the FEM in 1909, later he developed an efficient method for approximate problems [**22**]. His idea involved approximating the power function through known functions with unknown parameters.
The study of the finite element can be traced back to the works of Alexander Hrennikoff 1941 who created a frame method in which a flat, flexible medium is inter-

preted as a set of rails and girders. These pioneers share one important characteristic: the division of a continuous domain into a number of distinct subdomains, typically called elements.

Richard Currant In 1943, German mathematician, increased the probabilities of the Ritz method by introducing special linear functions defined via multiple-definition linear approximation in subareas [**12**] and using the finite element model of the procedure to reduce the potential energy of the torsion strain function using values Grid point as unknown parameters.

Using digital computers, 1950, solving a large number of equations simultaneously became possible, and the first published paper using the word 'Finite Element Method' was in 1960. Ray W. Clough. Zienkiewicz and Chung wrote their first book on 'Operation Unique Elements' in 1967. Therefor FEM was applied to a variety of engineering issues using FEM software packages (ABAQUS, NASTRAN, ANSYS, etc.).

In 1980 an algorithm was developed for electromagnetic, fluid flow, and thermal analysis applications using the FEM. Then engineers can analyzed methods to manage vibration and extend the use of diversity and to accelerate space structures using a finite and other methods. Trends to overcome additive solution to fluid flow are closely related to structural reactions and biomechanical problems, where a higher degree of accuracy was observed [**13**].

## 1.2 Advantages and disadvantages of the FEM

An important advantage of the FEM is the easily managing of the complex geometry, so, it can give a good approximations of a variety of engineering problems in the field of solid mechanics, fluid, dynamics, electrostatic problems, and heat problems. In addition, the FEM can manage dynamic constraints where an undetermined structure can be resolved.

However, the FEM has disadvantages, for example it just obtain a 'approximate' solution. As well as, a general closed-form solution that would allow a system response to a change in different parameters to be examined is not generated [**10**].

## 1.3 Preliminarily

**Definition 1.1.** [**9**] *$L_p-$spaces, For $p \in [1, \infty)$,*

$$L_p(\Omega) := \{v : \Omega \to \mathbb{R};\ \int_\Omega |v(x)|^p dx < \infty\}. \tag{1.1}$$

$$||v||_{l_p(\Omega)} := \left(\int_\Omega |v(x)|^p dx\right)^{\frac{1}{p}}. \tag{1.2}$$

*For $p = \infty$,*

$$L_\infty(\Omega) := \{v : \Omega \to \mathbb{R};\ |v(x)| < \infty\ a.e.\}.$$

$$||v||_{L_\infty(\Omega)} := \inf\{k > 0, |v(x)| \leq k\ a.e.\}.$$

*The integral (2.2) is called Lebesgue integral and 'a.e.' means 'almost every where'* [**16**]*, i.e. $\forall x \in \Omega \backslash \mathbb{N}$, for null sets $\mathbb{N}$.*

Important properties

1. The space $(L_p(\Omega), ||.||_{L_p})$ is Banach space for $p \in \mathbb{N}$.
2. The space $(L_2(\Omega), \langle . \rangle_{L_2(\Omega)})$ is a Hilbert space [**2**], where the inner product in $L_2$ is defined as
$$\langle \varphi, \psi \rangle_{L_2(\Omega)} = \int_\Omega \varphi(x)\psi(x)dx.$$

**Notation 1.2.** [**9**] *The space $C_c^\infty$ denoted the infinitely differentiable space functions $\psi : \Omega \to \mathbb{R}$ with compact support in $\Omega$.*

**Definition 1.3.** [**9**] *Assume a function $u \in C^1(\Omega)$. If $\psi \in C_c^\infty$ we give the formula of integration by parts*

$$\int_\Omega u\psi_{x_i} dx = -\int_\Omega u_{x_i}\psi dx, \tag{1.3}$$

*there is no boundary term since $\psi$ is with compact support in $\Omega$. If $k$ is a positive integer, $u \in C^k(\Omega)$, and $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d)$ is a multi-index of order $|\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_d = k$, then*

$$\int_\Omega u D^\alpha \psi dx = (-1)^{|\alpha|} \int_\Omega D^\alpha \psi dx,\ \forall \psi \in C_c^k(\Omega),$$

*where*

$$D^\alpha \psi = \left(\frac{\partial^{\alpha_1}}{\partial x_1^{\alpha_1}} \cdots \frac{\partial^{\alpha_n}}{\partial x_n^{\alpha_n}}\right)(\psi).$$

**Remark 1.4.** *Given a domain $\Omega$, a set of locally integrable functions is defined by* [**4**]

$$L_{loc}^1(\Omega) := \{g : g \in L^1(E), \forall\ compact\ (E) \subset interior(\Omega)\}.$$

### 1.3.1 Weak derivative

Suppose $f, g \in L^1_{loc}(\Omega)$ and $\alpha$ is a multi-index, we say that $g$ is the $\alpha^{\text{th}}$- weak partial derivative of $f$, written $g = D^\alpha f$, if [**9, 4**]

$$\int_\Omega f D^\alpha \psi dx = (-1)^{|\alpha|} \int_\Omega g\psi dx, \ \forall \psi \in C_c^\infty(\Omega),$$

or equivalently

$$\langle f, D^\alpha \psi \rangle_{L_2}(\Omega) = (-1)^{|\alpha|} \langle g, \psi \rangle_{L_2(\Omega)}, \ \forall \psi \in C_c^\infty(\Omega).$$

**Definition 1.5.** *Given a function $g \in L^1_{loc(\Omega)}$ we say that $h \in L^1_{loc(\Omega)}$ has a weak derivative $D^\alpha h$ if*

$$\int_\Omega h(x) D^\alpha \psi(x) dx = (-1)^{|\alpha|} \int_\Omega D^\alpha h(x) \psi(x) dx, \ \forall \psi \in C_c^\infty(\Omega).$$

**Remark 1.6.**

- *If a locally integrable function has a weak derivative, then it is unique, i.e., if $v = D^\alpha u \in L^1_{loc(\Omega)}$ and $\tilde{v} = D^\alpha u \in L^1_{loc(\Omega)}$ both are weak partial derivatives of $u$, then $v = \tilde{v}$ a.e.,* [**20**].
- *Consistency in the definition: If $u \in C^1(\Omega) \cap C(\bar{\Omega})$, then the weak derivative matches the classical derivative,* [**2**].

**Definition 1.7.** [**14**] *Let $k$ be a non-negative integer, and let $\psi \in L^1_{loc}$ be assumed to have a weak derivative $D^\alpha(\psi)$ for all $|\alpha| \leq k$, .Define the sobolev space $W_p^k$*

$$W_p^k := \{\psi \in L^1_{loc} : ||\psi||_{W_p^k} < \infty\},$$

*where for $1 \leq p < \infty$,*

$$||\psi||_{W_p^k(\Omega)} := (\sum_{|\alpha| \leq k} ||D^\alpha \psi||^p_{L_p(\Omega)})^{\frac{1}{p}},$$

*and for $p = \infty$,*

$$||\psi||_{W_\infty^k(\Omega)} := \max_{|\alpha| \leq k} ||D^\alpha \psi||_{L_\infty(\Omega)}.$$

**Remark 1.8.**

- *If $p = 2$ we usually write*

$$W_2^k(\Omega) = H^k(\Omega) = \{\psi \in L_2(\Omega) : \sum_{|\alpha| \leq k} D^\alpha \psi \in L_2(\Omega)\}, \ k = 0, 1, \cdots$$

*We use the letter $H$ and $H^k(\Omega)$ to denote for Hilbert space with inner product*

$$\langle u, v \rangle_{W_2^k} = \sum_{|\alpha| \leq k} \langle D^\alpha u, D^\alpha v \rangle.$$

- *The special case when $k = 1$ and $p = 2$ the space is*

$$H^1 = \{\psi \in L_2 : \frac{\partial \psi}{\partial x_i} \in L_2, i = 1, 2, \cdots, n\}. \tag{1.4}$$

*Note that*

$$||\psi||_{H^1(\Omega)} = (||\psi||^2_{L^2(\Omega)} + ||D\psi||^2_{L^2(\Omega)})^{\frac{1}{2}}.$$

**Definition 1.9.** [2] *The Sobolev space $H^k_0$ is the completion of the $C^\infty_c$ with respect to the norm $|| \cdot ||_{H^k}$, i.e.,*

$$u \in H^k_0(\Omega) \Longleftrightarrow \exists v \in C^\infty_c(\Omega) \text{such that} \lim_{n\to\infty} ||u - v||_{H^k(\Omega)} = 0.$$

*Note that $H^k_0(\Omega)$ is a closed subspace of $H^k$. If the boundary $\Gamma$ is $C^1$, then it is assumed that $v \in C(\bar{\Omega}) \cap H^k_0(\Omega)$ implies that $v(x) = 0$ for all $x \in \Gamma$. Finally, the special Sobolev space $H^1_0$ is defined as the closure of $C^\infty_0$ in $H^1(\Omega)$, so*

$$H^1_0 = \{u \in H^1(\Omega) : u \mid \Gamma = 0\}.$$

**Definition 1.10.** [13] *Let $(V, (\cdot, ))$ be an inner product space, if the associated normed linear space $(V, ||.||)$ is complete,then $(V, (, ))$ is called a Hilbert space.*

**Notation 1.11.** [13] *$H^1_0$ is a Hilbert space have the same norm and same inner product as of $H^1$.*

**Theorem 1.12.** [14] *The Sobolev space $H^k_p \equiv W^k_p$ with regard to the norm $||.||_{H^k_p}$ is a Banach space.*

**Notation 1.13.** [14] *With the Hilbert space $V$, the dual space $V'$ can be defined as the space of all linear functional $L(v)$. The linear functional $L(v)$ is bounded if $L(v) \leq C||v||_V, \forall v \in V$.*

**Lemma 1.14.** [15]*(Poincaré $-$ Frederic′s inequality). Let $\Omega$ be a bounded set of $\mathbb{R}^n$ for any $n$, then a constant $C_\Omega$ exists such that*

$$||u||_{L^2(\Omega)} \leq C_\Omega ||u||_{H^1(\Omega)}, \ \forall u \in H^1_0(\Omega).$$

## 1.4 Classification of the PDE

Three distinct families of partial differential equations are known: elliptical, parabolic and hyperbolic equations, where are classification is based on suitable unique computational methods. The general form of the linear second-order PDE is **[15, 17]**,

$$A(x,y)U_{xx}+B(x,y)U_{xy}+C(x,y)U_{yy}+D(x,y)U_x+E(x,y)U_y+F(x,y)U=G(x,y) \tag{1.5}$$

The classification depends on the sign of the discriminant, $\triangle = B^2 - 4AC$. In particular, equation (2.4) is called :

1. Elliptic equation, if $\triangle < 0$.
   As a standard example the Poisson equation

   $$\nabla^2 U = h(x,y) \text{ or } U_{xx}+U_{yy}=h.$$

   If $h=0$ it is called Laplace equation.

   $$\nabla^2 U = 0 \text{ or } U_{xx}+U_{yy}=0.$$

2. Parabolic equation, if $\triangle = 0$. As an example the Heat equation

   $$U_t = \alpha^2 U_{xx}.$$

3. Hyperbolic equation, if $\triangle > 0$. The Wave equation is an example of such equation

   $$U_{tt} - \alpha^2 U_{xx} = 0.$$

**Remark 1.15.** **[23]** *Important elliptic two dimension partial differential equations:*

$$u_{xx}+u_{yy}=0 \textit{ (Laplace Equation)}$$

$$-(u_{xx}+u_{yy})=h(x,y) \textit{ (Poission Equation)}$$

$$-(u_{xx}+u_{yy})+au=h \textit{ (General Helmholtz Equation)}$$

$$u_{xxxx}+2u_{xxyy}+u_{yyyy}=0 \textit{ (Bi-harmonic Equation)}$$

**Remark 1.16.** *The usual three type of boundary conditions:*

- *Dirichlet boundary condition: The solution is known at the boundary of the domain.*
- *Neumann boundary condition: The derivative of the solution is known at the boundary of the domain.*

- *Robin boundary condition: A mixed of* 1 *and* 2.

*Next, we shall discus the numerical technique of the finite element method for solving partial differential equations. Two major steps of the FEM that are the variational formulation and the discretization.*

# 1.5 Finite Element Method

## 1.5.1 Notations and definitions

Below are basic notations that will be used later in the finite element technique explanation [**3**].

- Element domain: A bounded closed set $k \subseteq I\!R^n$ where the PDE is associated.
- $V^L$ is the space of continuous piecewise linear polynomial
- $V_h^L$ is a finite subspace of $V^L$ on the partition $k_h : a = x_0 < x_1 < \cdots < x_n < x_{n+1} = b$.
- $\{\varphi_i\}_{i=0}^{n+1} = \{\varphi_0, \varphi_1, \cdots, \varphi_{n+1}\}$ is used for the set of nodal variables, where $\{\varphi\}_{i=0}^{n+1}$ are basis for $V_h^L$ satisfying

$$\varphi_i(x_j) = \begin{cases} 1, & \textit{if } i = j, \\ 0, & \textit{if } i \neq j, \end{cases}$$

  where $i, j = 0, 1, \cdots, n+1$.

Basis function $\varphi_i$ are continuous, piecewise linear and take the unite value at the node $x_i$, and zero at all other nodes. Note that $\varphi_i$ are known as the hat functions, because of their shape, see figure 1. The explicit expression for the hat functions are given by

$$\varphi_i(x) = \begin{cases} \frac{x - x_{i-1}}{h_i}, & x_{i-1} \leq x < x_i, \\ \frac{x_{i+1} - x}{h_{i+1}}, & x_i \leq x \leq x_{i+1}, \\ 0, & \text{elsewhere}, \end{cases}$$

where $h_i = x_i - x_{i-1}$.

Now Any function $u \in V_h^L$ can be written as a linear combination of the hat function $\{\varphi_i\}_{i=0}^{n+1}$ and corresponding coefficient $\{\xi_i\}_{i=0}^{n+1}$

$$u(x) = \sum_{i=0}^{n+1} \xi_i \varphi_i(x)$$

where $\xi_i = u(x_i), i = 0, \cdots, n+1$, are the value of the unknown at the nodal value $x_i$ to be determined. Finally, the notation $(k_h, V_h^L, \varphi)$ is known as the finite element triple.

**Remark 1.17.**

- *$V_h^L$ is a subspace of $V^L$ consisting of linear functions spanned by $\{\varphi_i\}_{i=0}^{n+1}$.*
- *The interval $[x_{i-1}, x_{i+1}]$ is called the support of the function $\varphi_i$, $i = 1, \cdots, n$.*
- *The exception of $\varphi_0$ and $\varphi_{n+1}$ at the left-end node $x_0$ and the right-end node $x_{n+1}$ with support only on one subinterval,i.e.,*

$$\varphi_0(x) = \begin{cases} \frac{x_1 - x}{h_1}, & x_0 \le x < x_1, \\ 0, & \textit{elsewhere}, \end{cases}$$

$$\varphi_{n+1}(x) = \begin{cases} \frac{x - x_n}{h_{n+1}}, & x_n \le x < x_{n+1}, \\ 0, & \textit{elsewhere}. \end{cases}$$

### 1.5.2 How the FEM works

We illustrate the basic steps of the finite element method below:

1. **Discretization:**

The solution area is to be divided into finite elements. Description of mesh consists from several main matrices, including nodal coordinates and element conductivities.

2. **Interpolation:**

Domain variables on the element are interpolated using interpolation functions. The choice of polynomials as interpolation functions is often used. However, the number of nodes assigned to the element decides the degree of the polynomial.

3. **Variational formulation:**

The matrix equation for the finite element should be established which associates the nodal values of the unknown function with other parameters. For this aim, different approaches can be used, the most suitable is: the variational formulation.

4. **Element equations:**

To find the global equation system for the whole solution area we must assemble all the equations of the elements. In other words, we must sum the local element equation for all elements used for discretization. Element connections are used for assembly process. Before solving, the boundary conditions (which are not computed in the element equations) must be imposed.

5. **System of global equations:**
The global equations for the finite element typically sparse, metric and positive definite. Both direct and iterative methods can be used to solve the resulted system. The nodal values of the sought function are produced as a result of the solution.

### 1.5.3 Stiffness and Mass matrices

To ensure an approximate solution of the resulting differential equation, matrices like stiffness and mass matrices that represent a system of linear equations must be solved.

**Stiffness matrix**
Consider the matrix $A = [a_{ij}]$, where $a_{ij} = \int_\Omega \varphi_i^{'}\varphi_j^{'}\, dx$ and where $\varphi_i$ and $\varphi_j$ are the linear basis functions defined before. If $|j-i| > 1$, then $\varphi_i^{'}\varphi_j^{'} = 0$ and so $a_{ij} = 0$. If $|j-i| \leq 1$, then one of the following cases holds:

- $j - i = -1 \Rightarrow j = i - 1$,
- $j - i = 0 \Rightarrow j = i$,
- $j - i = 1 \Rightarrow j = i + 1$.

The case $j = i - 1$

$$a_{i,i-1} = \int_\Omega \varphi_{i-1}^{'}\varphi_i^{'}\, dx = \int_{x_{i-1}}^{x_i} -\frac{1}{h_i}\frac{1}{h_i}\, dx = -\frac{1}{h_i}$$

while $j = i$

$$a_{i,i} = \int_\Omega \varphi_i^{'}\varphi_i^{'}\, dx = \int_{x_{i-1}}^{x_i} \frac{1}{h_i}\frac{1}{h_i}\, dx + \int_{x_i}^{x_{i+1}} \left(-\frac{1}{h_{i+1}}\right)\left(-\frac{1}{h_{i+1}}\right)\, dx = \frac{1}{h_i} + \frac{1}{h_{i+1}}$$

and $j = i + 1$

$$a_{i,i+1} = \int_\Omega \varphi_{i+1}^{'}\varphi_i^{'}\, dx = \int_{x_i}^{x_{i+1}} \left(\frac{1}{h_{i+1}}\right)\left(-\frac{1}{h_{i+1}}\right)\, dx = -\frac{1}{h_{i+1}}.$$

Thus, the stiffness matrix for non-uniform mesh is given by

$$A = \begin{pmatrix} \frac{1}{h_1} & -\frac{1}{h_2} & 0 & 0 & & 0 \\ -\frac{1}{h_2} & \frac{1}{h_1} + \frac{1}{h_2} & -\frac{1}{h_3} & 0 & \cdots & 0 \\ 0 & -\frac{1}{h_3} & \frac{1}{h_2} + \frac{1}{h_3} & -\frac{1}{h_4} & & \vdots \\ & 0 & -\frac{1}{h_4} & & \ddots & \\ \vdots & & \ddots & & \ddots & 0 \\ & & & 0 & \frac{1}{h_{n-1}} + \frac{1}{h_n} & -\frac{1}{h_{n+1}} \\ 0 & \cdots & & 0 & -\frac{1}{h_{n+1}} & \frac{1}{h_{n+1}} \end{pmatrix}.$$

If $h_1 = h_2 = \cdots = h_{n+1}$ uniform mesh, then $A$ becomes

$$A = \frac{1}{h} \begin{pmatrix} 1 & -1 & 0 & & & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & & \vdots \\ \vdots & & \ddots & & \ddots & 0 \\ \vdots & & & & \ddots & -1 \\ 0 & \cdots & & 0 & -1 & 1 \end{pmatrix}.$$

**Mass matrix**

Consider the matrix $M = [m_{ij}]$, where $m_{ij} = \int_\Omega \varphi_i \varphi_j \, dx$ and where $\varphi_i$ and $\varphi_j$ are the linear basis functions defined before.

Now, $m_{ij} = 0$ except for $j = i-1, i, i+1$. accordingly we have the following three cases:

- j=i-1,

$$m_{i,i-1} = \int_\Omega \varphi_{i-1}\varphi_i \, dx = \int_{x_{i-1}}^{x_i} \varphi_{i-1}\varphi_i \, dx = \frac{h_i}{6}.$$

- j=i,

$$m_{i,i} = \int_\Omega \varphi_i\varphi_i \, dx = \int_{x_{i-1}}^{x_i} \varphi_i\varphi_i \, dx + \int_{x_i}^{x_{i+1}} \varphi_i\varphi_i \, dx = \frac{h_i + h_{i+1}}{6}.$$

- j=i+1,

$$m_{i,i+1} = \int_\Omega \varphi_{i+1}\varphi_i \, dx = \int_{x_i}^{x_{i+1}} \varphi_{i+1}\varphi_i \, dx = \frac{h_{i+1}}{6}.$$

Hence, the mass matrix for non-uniform mesh is given by:

$$M = \begin{pmatrix} \frac{h_1}{3} & \frac{h_2}{6} & 0 & & & 0 \\ \frac{h_2}{6} & \frac{h_1+h_2}{6} & \frac{h_2}{6} & 0 & \cdots & 0 \\ 0 & \frac{h_2}{6} & \ddots & \ddots & & 0 \\ 0 & & & & \cdots & \vdots \\ \vdots & & & & \frac{h_{n-1}+h_n}{6} & \frac{h_{n+1}}{6} \\ 0 & \cdots & & 0 & \frac{h_{n+1}}{6} & \frac{h_{n+1}}{3} \end{pmatrix}.$$

With uniform mesh the mass matrix $M$ is:

$$A = \frac{h}{6} \begin{pmatrix} 2 & 1 & 0 & \cdots & 0 \\ 1 & 4 & 1 & & 0 \\ 0 & 1 & 4 & & \vdots \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & 0 & 1 & 2 \end{pmatrix}.$$

### 1.5.4 Examples

**Example 1.18.** *Consider the following boundary value problem*

$$\begin{aligned} -u^{''} &= f, \quad x \in (0,1), \\ u(0) &= u(1) = 0. \end{aligned} \tag{1.6}$$

*For general $f$, the exact solution of this problem depends on the choice of the function $f$. For example,with $f = 1$ one can find that $u = \frac{x(1-x)}{2}$. However, it may be difficult to find $u$ with analytical techniques for some other choices of $f$. We will consider this BVP as a good model for studying the numerical techniques introduced by the FEM.*

***Variational formulation*** *Multiply* (1.6) *by a test function $v$, integrate over $\Omega = (0,1)$, and use that $v(0) = v(1) = 0$ to get*

$$\begin{aligned} \int_0^1 fv\,dx &= -\int_0^1 u^{''} v\,dx \\ &= \int_0^1 u^{'} v^{'}\,dx - u^{'}(1)v(1) + u^{'}(0)v(0) \\ &= \int_0^1 u^{'} v^{'}\,dx. \end{aligned}$$

*The statement of the variational formulation of* (1.6) *will be:*
*Find* $u \in H_0^1([0,1]) = \{w : ||w'|| < \infty, ||w|| < \infty, \ \textit{and } w(0) = w(1) = 0\}$ *such that*

$$\int_0^1 u^{'}v^{'}\,dx = \int_0^1 fv\,dx, \ \ \forall v \in H_0^1([0,1]). \tag{1.7}$$

**The FEM discretization.** Introducing the vector space, $V_h^L$, of continuous, piecewise linear functions on the partition $0 = x_0 < x_1 < \cdots < x_n < x_{n+1} = 1$, of $[0,1]$. We state the $cG(1)$ (see the remark below) method as the following discrete counterpart of (1.7) : Find $U(x) \in V_h^L$, such that

$$\int_0^1 U^{'}v^{'}\,dx = \int 0^1 fv\,dx \forall x \in V_h^L. \tag{1.8}$$

**Remark 1.19.** *In* $cG(1)$ *expressing the fact that this FEM is based on continuous, piecewise linear approximation. The letter* c *stands for continuous and* G *stands for Galerkin , and the number* 1 *stands for linear. Boris Grigorievich Galerkin* $(1871-1945)$ *was a Russian mathematician who made pioneering contributions to the field of numerical solution of differential equations. The Galerkin method is the method of rewriting the differential equation in variational form and discretizing it. A FEM, is a Galerkin method that utilises piecewise polynomials as approximating functions.*

We now seek a solution, $U(x)$, to (1.8) expressed of the hat function $\{\varphi_i\}_{i=0}^{n+1} \subset V_h^L$.
In other words we propose

$$U(x) = \sum_{j=0}^{n+1} \xi_j \varphi_j(x) \tag{1.9}$$

and search for the coefficients vector

$$\xi = \begin{pmatrix} \xi_0 \\ \xi_1 \\ \vdots \\ \xi_{n+1} \end{pmatrix} = \begin{pmatrix} U(x_0) \\ U(x_1) \\ \vdots \\ U(x_{n+1}) \end{pmatrix}$$

of the nodal values $U(x)$ in such a way that (1.8) is satisfied. Using the fact that $\xi_0 = U(x_0) = U(0) = 0$ and $\xi_{n+1} = U(x_{n+1}) = U(1) = 0$, then $U$ can be written as

$$U(x) = \sum_{j=1}^{n} \xi_j \varphi_j(x).$$

To derive the linear system of equations, we substitute (1.9) into (1.8),

$$\int_0^1 \sum_{j=0}^{n+1} \xi_j \varphi_j^{'}(x) v^{'} \, dx = \int_0^1 fv \, dx, \quad \forall v \in V_h^L. \tag{1.10}$$

Since $\{\varphi_i\}_{i=0}^{n+1} \subset V_h^L$ is a basis of $V_h^L$ then we take $v = \varphi_i$, so, (1.10) is equivalent to

$$\int_0^1 \sum_{j=1}^{n} \xi_j \varphi_j^{'}(x) \varphi_i^{'}(x) \, dx = \int_0^1 f\varphi_i \, dx, \quad i = 1, \cdots, n$$

$$\int_0^1 \xi_j \sum_{j=1}^{n} \varphi_j^{'}(x) \varphi_i^{'}(x) \, dx = \int_0^1 f\varphi_i \, dx, \quad i = 1, \cdots, n. \tag{1.11}$$

This is a quadratic system of $n$ linear equation and $n$ unknowns. Introducing the notations

$$\begin{aligned} a_{i,j} &= \int_0^1 a\varphi_j'\varphi_i' dx, \\ b_i &= \int_0^1 f\varphi_i dx, \end{aligned}$$

Then, in matrix form, the system (1.11) is read as $A\xi = b$, where

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{nn} \end{bmatrix}$$

is the stiffness matrix and

$$b = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}$$

is the load vector.

**Example 1.20.** *Consider the following boundary value problem with* $n = 3$

$$-u'' = 1, \quad x \in (0,1),$$

$$u(0) = u(1) = 0.$$

Solution:
*(a) The exact solution is* $u(x) = -\frac{x^2}{2} + \frac{x}{2}$.
*(b) the numerical solution:*
*Let* $v \in H_0^1$ *be a test function, multiply the equation with* $v$ *then integrating by parts over* $[0, 1]$

$$\begin{aligned} \int_0^1 -u''v\,dx &= \int_0^1 v\,dx \\ -u'v'|_0^1 + \int_0^1 u'v'\,dx &= \int_0^1 v\,dx \end{aligned}$$

*Then,*

$$\begin{aligned} -u(1)v(1) + u(0)v(0) + \int_0^1 u'v'\,dx &= \int_0^1 v\,dx \\ \int_0^1 u'v'\,dx &= \int_0^1 v\,dx \end{aligned}$$

*Now, let* $U \in V_h^L$ *with* $U = \sum_{j=0}^3 \xi_j \varphi_j$ *and* $v = \varphi_i$, *then*

$$\begin{aligned} \int_0^1 \sum_{j=0}^3 \xi_j \varphi_j' \varphi_i'\,dx &= \int_0^1 \varphi_i\,dx \\ \sum_{j=0}^3 \xi_j \int_0^1 \varphi_j' \varphi_i'\,dx &= \int_0^1 \varphi_i\,dx \qquad (1.12) \end{aligned}$$

*we have* $h = \frac{1-0}{3} = \frac{1}{3}$ *and we know that* $U(x_j) = \xi_j,\ j = 0, 1, 2, 3$. *Hence* $U(0) = U(x_0) = \xi_0 = 0$, $U(1) = U(x_3) = \xi_3 = 0$, *so,* (1.12) *can be written as*

$$\sum_{j=1}^2 \xi_j \int_0^1 \varphi_j' \varphi_i'\,dx = \int_0^1 \varphi_i\,dx, \quad i = 1, 2. \qquad (1.13)$$

*Note that,*

$$\varphi_1 = \begin{cases} \frac{x-0}{\frac{1}{3}}, & 0 \le x < \frac{1}{3}, \\ \frac{\frac{2}{3}-x}{\frac{1}{3}}, & \frac{1}{3} \le x < \frac{2}{3}, \\ 0, & \textit{elsewhere.} \end{cases} \quad \varphi_2 = \begin{cases} \frac{x-\frac{1}{3}}{\frac{1}{3}}, & \frac{1}{3} \le x < \frac{2}{3}, \\ \frac{1-x}{\frac{1}{3}}, & \frac{2}{3} \le x < 1, \\ 0, & \textit{elsewhere.} \end{cases}$$

*After computation the elements integral in* (1.13) *we end up with*

$$6\xi_1 - 3\xi_2 = \frac{1}{3}, \quad -3\xi_1 + 6\xi_2 = \frac{1}{3}.$$

*Hence,* $\xi_1 = \xi_2 = \frac{1}{9}$.
*This implies,*

$$\begin{pmatrix} \xi_0 \\ \xi_1 \\ \xi_2 \\ \xi_3 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{9} \\ \frac{1}{9} \\ 0 \end{pmatrix}.$$

*Therefore,* $U(x) = \frac{1}{9}\varphi_1(x) + \frac{1}{9}\varphi_2(x)$ *is the approximation solution using the FEM with* $n = 3$.

*Chapter 2*

# Poisson equation

## 2.1 Poisson Applications

Poisson equation is the property of the class of elliptic equations and has numerous applications in physics and mechanics. These include,[**8**]
**Electrostatics.**
Let $E(x)$ be the electric field in a volume $\Omega$ containing charges of density $\rho(x)$ and enclosed by a perfectly conducting surface $\Gamma$. According to Maxwell equations describing electromagnetic problem:

$$\nabla \cdot E = \rho \,\text{in}\, \Omega,$$

where $E(x)$ is a conservative field. From Faraday's law,

$$\nabla \times E = 0,$$

it follows that there exist a scalar electric potential $\varphi$, such that

$$E = \nabla\varphi.$$

This leads to the Poisson equation

$$\nabla \cdot \nabla\varphi = \triangle\varphi = \rho$$

with a Dirichlet boundary condition $\varphi = c$ on $\Gamma$, where $c$ is a constant.

**Fluid mechanics.**
The rotation-free fluid flow is a conservative field and satisfies

$$\nabla \times u = 0,$$

where $u$ is the velocity field. It follows that there exist a scalar velocity potential $\varphi$ such that

$$u = \nabla \psi.$$

For incompressible fluid we have $\nabla \cdot u = 0$, and we obtain the Laplace equation for the potential of rotation-free incompressible flow

$$\nabla \cdot \nabla \varphi = \triangle \varphi = 0.$$

At a solid boundary, the normal velocity is zero, which translates to a homogeneous Neumann boundary condition for the potential.

**Statistical physics.**

In this application, we consider the random motion of particles inside a container $\Omega$. The particles move until they hit the boundary where they stop. We assume that the boundary $\Gamma$ of $\Omega$ is partitioned into two parts, see Fig

$$\Gamma = \Gamma_1 \cup \Gamma_2, \Gamma_1 \cap \Gamma_2 = \phi.$$

Let $u(x)$ be the probability that a particle starting at $x \in \Omega$ winds up stopping at some point on $\Gamma_1$, so that $u(x) = 0$ means it never happens, and $u(x) = 1$ means that it is certain to happen. It turns out that $u$ follows Laplace equation

$$\triangle u = 0 \text{ in } \Omega,$$

with boundary conditions

$$u|_{\Gamma_1} = 1 \ \text{ and } u|_{\Gamma_2} = 0.$$

The solution of this boundary value problem, as expected, is not continuous on the boundary.

## 2.2 One dimensional Poisson equation

### 2.2.1 Modeling and Variational Formulation

Consider the stationary reaction-diffusion process involving a single substance, which has the following mathematical model

$$\begin{aligned} -(au')' &= f, \ \ \alpha < x < \beta, \\ a(\alpha)u'(\alpha) &= b(\alpha)(u(\alpha) - g_D(\alpha)) + g_N(\alpha), \\ -a(\beta)u'(\beta) &= b(\beta)(u(\beta) - g_D(\beta)) + g_N(\beta), \end{aligned} \tag{2.1}$$

where the unknown $u(x)$ stands the concentration of the substance, and the other functions:
$a(x)$ : diffusion coefficient $\quad a(x) > 0$,
$f(x)$ : source function
$b(\alpha), b(\beta)$ : permeability at the end points $\quad b \geq 0$,
$g_D(\alpha), g_D(\beta)$ : ambient concentration
$g_N(\alpha), g_N(\beta)$ : externally induced flux through the boundary

First, turning to the Boundary conditions in (2.1), we will consider the cases for which a mixed boundary conditions are representing a mathematical model of the physical fact that the outward flux is proportional to the concentration difference between the domain boundary and the surrounding environment, i.e, $g_N(\alpha) = g_N(\beta) = 0$.
The following three special cases will be considered:

- Dirichlet boundary conditions:
  This boundary condition physically corresponds to the case of a vary high permeability, i.e, $b \to \infty$, implying that the concentration at the boundary adapts to the ambient concentration $u = g_D$.

- Homogeneous Neunmman boundary condition:
  This boundary condition physically corresponds to the case of an impermeable boundary, i.e, one for which $b = 0$ and $g_N(\alpha) = g_N(\beta) = 0$, implying zero flux through the boundary, that is $u'(\alpha) = u'(\beta) = 0$.

- Inhomogeneous Neunmman boundary condition:
  We can also imagine a case where we externally control the flux through the boundary. This case can be modelled by choosing $b = 0$ and $g_N \neq 0$.
  This boundary condition prescribes the flux through the boundary

$$a(\alpha)u'(\alpha) = g_N(\alpha), \quad -a(\beta)u'(\beta) = g_N(\beta).$$

The derivation of variational formulation of (2.1) is explained in the following steps, but first we define the space

$$H^1([\alpha,\beta]) := \{v(x) : \int_\alpha^\beta v^2(x)\, dx < \infty, \int_\alpha^\beta (v')^2(x)\, dx < \infty\},$$

that will be used later.

- We multiply the differential equation (2.1) by test function $v(x) \in H^1([\alpha, \beta])$.
- Integrate both sides over $[\alpha, \beta]$

  $$-\int_{\alpha}^{\beta} (au')'v \, dx = \int_{\alpha}^{\beta} fv \, dx,$$

  then we use integration by parts,

  $$[-(au')v]_{\alpha}^{\beta} + \int_{\alpha}^{\beta} au'v' \, dx = \int_{\alpha}^{\beta} fv \, dx,$$

- Make use the boundary conditions in (2.1)

  $$\begin{aligned} a(\alpha)u'(\alpha) &= b(\alpha)(u(\alpha) - g_D(\alpha)) + g_N(\alpha), \\ -a(\beta)u'(\beta) &= b(\beta)(u(\beta) - g_D(\beta)) + g_N(\beta), \end{aligned}$$

  to obtain:

  $$\begin{aligned} b(\beta)u(\beta)v(\beta) + b(\alpha)u(\alpha)v(\alpha) + \int_{\alpha}^{\beta} au'v' \, dx &= (b(\beta)g_D(\beta) - g_N(\beta))v(\beta) \\ &+(b(\alpha)g_D(\alpha) - g_N(\alpha))v(\alpha) + \int_{\alpha}^{\beta} fv \, dx. \end{aligned}$$

- Finally, the statement of the variational formulation of (2.1) becomes:
  Find $u(x) \in H^1([\alpha, \beta])$, such that

  $$\begin{aligned} b(\beta)u(\beta)v(\beta) + b(\alpha)u(\alpha)v(\alpha) + \int_{\alpha}^{\beta} au'v'dx &= (b(\beta)g_D(\beta) - g_N(\beta))v(\beta) \quad (2.2) \\ &+(b(\alpha)g_D(\alpha) - g_N(\alpha))v(\alpha) + \int_{\alpha}^{\beta} fvdx, \;\; \forall v \in H^1([\alpha, \beta]) \end{aligned}$$

### 2.2.2 Discretizaion

Let $V_h^L$ be the vector space of continuous, piecewise linear functions on the partition of $[\alpha, \beta]$, $\alpha = x_1 < x_2 < \cdots < x_{N-1} < x_N = \beta$. We now state the $cG(1)$ method as the following discrete counterpart of (2.2) :
Find $U(x) \in V_h^L$, such that

$$\begin{aligned} b(x_N)U(x_N)v(x_N) + b(x_1)U(x_1)v(x_1) + \int_{x_1}^{x_N} aU'v'dx = (b(x_N)g_D(x_N) \quad (2.3) \\ -g_N(x_N))v(x_N) + (b(x_1)g_D(x_1) - g_N(x_1))v(x_1) + \int_{x_1}^{x_N} fvdx, \;\; \forall v \in V_h^L. \end{aligned}$$

**Assumption**

We express the solution $U(x)$, of (2.3) in terms of the basis $\{\varphi_i\}_{i=1}^N \subset V_h^L$ as defined before.

As a result, we seek a solution of the form

$$U(x) = \sum_{j=1}^{N} \xi_j \phi_j(x) \tag{2.4}$$

and follow the computations below to determine the coefficient vector,

$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_N \end{bmatrix} = \begin{bmatrix} U(x_1) \\ U(x_2) \\ \vdots \\ U(x_N) \end{bmatrix}$$

of nodal values of $U(x)$, in such a way that (2.3) is satisfied.

Now, substitute (2.4) into (2.3),

$$\begin{aligned} b(x_N)\xi(x_N)\upsilon(x_N) + b(x_1)\xi(x_1)\upsilon(x_1) + \sum_{j=1}^{N} \xi_j \int_{x_1}^{x_N} a\phi_j' \upsilon' dx = (b(x_N)g_D(x_N) \quad (2.5) \\ -g_N(x_N))\upsilon(x_N) + (b(x_1)g_D(x_1) - g_N(x_1))\upsilon(x_1) + \int_{x_1}^{x_N} f\upsilon dx, \quad \forall \upsilon \in V_h^L. \end{aligned}$$

Since $\{\varphi_i\}_{i=1}^N \subset V_h^L$ is a basis of $V_h^L$, then we can set $\upsilon = \varphi_i,\ i = 1, \cdots$ ,N, thus equation (2.5) becomes

$$\begin{aligned} b(x_N)\xi(x_N)\varphi_i(x_N) + b(x_1)\xi(x_1)\varphi_i(x_1) + \sum_{j=1}^{N} \xi_j \int_{x_1}^{x_N} a\varphi_j' \varphi_i' \, dx = (b(x_N)g_D(x_N)(2.6) \\ -g_N(x_N))\varphi_i(x_N) + (b(x_1)g_D(x_1) - g_N(x_1))\varphi_i(x_1) + \int_{x_1}^{x_N} f\varphi_i \, dx, \quad i = 1, 2, \cdots, N, \end{aligned}$$

which is a system of $N$ linear equations and $N$ unknowns. Using the notation

$$\begin{aligned} a_{i,j} &= \int_{x_1}^{x_N} a\varphi_j' \phi_i' \, dx, \\ b_i &= \int_{x_1}^{x_N} f\varphi_i \, dx, \end{aligned}$$

and noting that

$$\varphi_i(x_1) = \begin{cases} 1, & \text{if } i = 1, \\ 0, & \text{if } i \neq 1, \end{cases}$$

and

$$\varphi_i(x_N) = \begin{cases} 1, & \text{if } i = N, \\ 0, & \text{if } i \neq N, \end{cases}$$

we can write the system of equations (2.6) as a discrete system of linear equations as:

$$\begin{aligned} (b(x_1) + a_{1,1})\xi_1 + \cdots + a_{1,N}\xi_N &= b_1 + b(x_1)g_D(x_1) - g_N(x_1), \\ a_{2,1}\xi_1 + \cdots + a_{2,N}\xi_N &= b_2, \\ \vdots &= \vdots \\ a_{N-1,1}\xi_1 + \cdots + a_{N-1,N}\xi_N &= b_{N-1}, \\ a_{N,1}\xi_1 + \cdots + (a_{N,N} + b(x_N))\xi_N &= b_N + b(x_N)g_D(x_N) - g_N(x_N). \end{aligned} \tag{2.7}$$

In matrix form, this reads,

$$(A + R)\xi = \tilde{b} + rv,$$

where

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \cdots & a_{N,N} \end{bmatrix},$$

$R$ is the boundary contributions to the system matrix given by

$$R = \begin{bmatrix} b(x_1) & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & b(x_N) \end{bmatrix},$$

$\tilde{b}$ is the load vector

$$\tilde{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix},$$

and

$$rv = \begin{bmatrix} b(x_1)g_D(x_1) - g_N(x_1) \\ 0 \\ \vdots \\ 0 \\ b(x_N)g_D(x_N) - g_N(x_N) \end{bmatrix}$$

contains the boundary contributions to the right side of (2.7).

## 2.3 Two Dimensional Poisson equation

### 2.3.1 Modeling and Variational Formulation

Consider the stationary reaction-diffusion process involving a single substance, which the following mathematical model

$$\begin{aligned} -\nabla \cdot (a\nabla u) &= f, \quad x = (x_1, x_2) \in \Omega \subset \mathbb{R}^2 \\ -n \cdot (a\nabla u) &= b(u - g_D) + g_N, \quad x = (x_1, x_2) \in \Gamma = \partial\Omega, \end{aligned} \tag{2.8}$$

where the unknown function $u = u(x_1, x_2)$ denotes the concentration of the substance. Below we list the meaning of the functions appeared in the problem:
$a(x_1, x_2) : \Omega \to \mathbb{R}$ diffusion coefficient, $a(x_1, x_2) > 0$
$f(x_1, x_2) : \Omega \to \mathbb{R},$ source function
$b(x_1, x_2) : \partial\Omega \to \mathbb{R}$ permeability of the boundary, $b(x_1, x_2) \geq 0$
$g_D(x_1, x_2) : \partial\Omega \to \mathbb{R}$ ambient concentration
$g_N(x_1, x_2) : \partial\Omega \to \mathbb{R}$ externally induced flux through the boundary

We first consider the case $g_N = 0$ for all $x = (x_1, x_2) \in \partial\Omega$, for which Robin boundary conditions are a mathematical model of the physical fact, that is the flux through the boundary, $-n \cdot (a\nabla u) = -a\frac{\partial u}{\partial n}$ where $n(x) = (n_1(x_1, x_2), n_2(x_1, x_2))$ denotes the outward unit normal on $\partial\Omega$, is proportional to the concentration difference between the domain boundary and its surroundings. Note that, since $n$ is taken to be the outward unit normal, a positive sign corresponds to an outward flux. We have the following cases:

- Dirichlet boundary condition:
  This boundary condition physically corresponds to the case of very high permeability, i.e., $b \to +\infty$, implying that the concentration at the boundary adapts to the ambient concentration. $u = g_D$. (The special case $u = 0$, is referred to as a homogeneous Dirichlet boundary condition.)

- Homogeneous Neumann boundary condition:
  This boundary condition physically corresponds to the case of an impermeable boundary, i.e., one where $b = 0$ and $g_N = 0$, implying zero flux through the boundary: $-n \cdot (a\nabla u) = 0$.

- Inhomogeneous Neumann boundary condition:
  This boundary condition prescribes the flux through the boundary, which can be obtained by assuming $b = 0$, thus we get $-n \cdot (a\nabla u) = g_N$.

The derivation of variational formulation of (2.8) is explained in the following steps:

- Multiply the differential equation by a test function $v = v(x_1, x_2)$.

- Integrate both sides over $\Omega$

$$-\int\int_{\Omega} \nabla \cdot (a\nabla u)v \; dx_1 \; dx_2 = \int\int_{\Omega} fv \; dx_1 \; dx_2,$$

  that is

$$-\int\int_{\Omega} (\frac{\partial}{\partial x_1}(a\frac{\partial u}{\partial x_1}) + \frac{\partial}{\partial x_2}(a\frac{\partial u}{\partial x_2}))v \; dx_1 \; dx_2 = \int\int_{\Omega} fv \; dx_1 \; dx_2.$$

  Now, using integration by parts to obtain

$$-\int_{\partial\Omega} (a\frac{\partial u}{\partial x_1}n_1 + a\frac{\partial u}{\partial x_2}n_2)v \; ds + \int\int_{\Omega} (a\frac{\partial u}{\partial x_1}\frac{\partial v}{\partial x_1} + a\frac{\partial u}{\partial x_2}\frac{\partial v}{\partial x_2}) \; dx_1 \; dx_2 = \int\int_{\Omega} fvdx_1dx_2.$$

  In vector notations it takes the form:

$$-\int_{\partial\Omega} (n.(a\nabla u))v \; ds + \int\int_{\Omega} a\nabla u.\nabla v \; dx_1 \; dx_2 = \int\int_{\Omega} fv \; dx_1 \; dx_2.$$

- Use the boundary condition in (2.8)

$$-n \cdot (a\nabla u) = b(u - g_D) + g_N, \quad x = (x_1, x_2) \in \partial\Omega,$$

  to obtain

$$\int_{\partial\Omega} buv \; ds + \int\int_{\Omega} a\nabla u.\nabla v \; dx_1 \; dx_2 = \int_{\partial\Omega} (bg_D - g_N)v \; ds + \int\int_{\Omega} fv \; dx_1 \; dx_2.$$

- Finally, the variational formulation of (2.8) has the following statement:
  Find $u \in V$, such that
  $$\int_{\partial\Omega} buv\ ds + \int\int_{\Omega} a\nabla u.\nabla v\ dx_1\ dx_2 = \int_{\partial\Omega} (bg_D - g_N)v\ ds + \int\int_{\Omega} fv\ dx_1\ dx_2, \tag{2.9}$$
  for all $v \in V$ where $V$ denotes the vector space of functions $v = v(x_1, x_2)$ that are sufficiently regular for the integrals in (2.9) to exist.

### 2.3.2 Discretizaion

Introduce the vector space, $V_h$, of continuous piecewise linear functions on a triangulation $T_h = \{K_i\}_{i=1}^{ntri}$, where ntri denotes the number of triangles: the mesh, of $\Omega$ (which is assumed to have a polygonal boundary), with the corresponding set of nodes, $N_h = \{N_i\}_{i=1}^{nnodes}$, where nnode denotes the number of nodes in the triangulation. We now state the $cG(1)$ method as the following discrete counterpart of (2.9) :
Find $U \in V_h$, such that
$$\int_{\partial\Omega} bUv\ ds + \int\int_{\Omega} a\nabla U.\nabla v\ dx_1\ dx_2 = \int_{\partial\Omega} (bg_D - g_N)v\ ds + \int\int_{\Omega} fv\ dx_1\ dx_2, \tag{2.10}$$
for all $v \in V_h$.

**Ansatz**
We now set a solution, $U(x_1, x_2)$, to (2.10), expressed in terms of the basis (tent functions) $\{\varphi_i\}_{i=1}^{nnodes} \subset V_h$ where $\varphi_i(N_j) = \delta_{ij}, i, j = 1, \cdots ,$nnodes, and where $\delta_{ij}$ denotes the Kronecker delta function defined by

$$\delta_{ij} = \begin{cases} 1, & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

In other words, we set
$$U(x_1, x_2) = \sum_{j=1}^{nnodes} \xi_j \varphi_j(x_1, x_2) \tag{2.11}$$
and seek to determine the coefficient vector,
$$\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_{nnodes} \end{bmatrix} = \begin{bmatrix} U(N_1) \\ U(N_2) \\ \vdots \\ U(N_{nnodes}) \end{bmatrix}$$

of nodal values of $U(x_1, x_2)$, in such a way that (2.10) is satisfied.

**Construction of discrete system of linear equations**

We substitute (2.11) into (2.10),

$$\begin{aligned}\sum_{j=1}^{nnodes} \xi_j \{ \int_{\partial\Omega} b\varphi_j v ds + \int\int_{\Omega} a \nabla\varphi_j \cdot \nabla v dx_1 dx_2 \} = \int_{\partial\Omega} (bg_D - g_N) v ds \qquad (2.12)\\ + \int\int_{\Omega} f v dx_1 dx_2, \text{for all } v \in V_h.\end{aligned}$$

Since $\{\varphi_i\}_{i=1}^{nnodes} \subset V_h$ is a basis of $V_h$, then we can assume $v = \varphi_i,\ i = 1, \cdots$ , nnodes, thus (2.12) is equivalent to

$$\begin{aligned}\sum_{j=1}^{nnodes} \xi_j \{ \int_{\partial\Omega} b\varphi_j \varphi_i ds + \int\int_{\Omega} a \nabla\varphi_j \cdot \nabla \varphi_i dx_1 dx_2 \} = \int_{\partial\Omega} (g_D - g_N) \varphi_i \ ds \qquad (2.13)\\ + \int\int_{\Omega} f \varphi_i dx_1 dx_2, \ \ i = 1, \cdots, \text{nnodes},\end{aligned}$$

which is system of nnodes linear equations and nnodes unknowns. Introducing the notation

$$\begin{aligned} r_{i,j} &= \int_{\partial\Omega} b\varphi_j \varphi_i ds, \\ a_{i,j} &= \int\int_{\Omega} a \nabla\varphi_j . \nabla\varphi_i dx_1 dx_2, \\ rv_i &= \int_{\partial\Omega} (bg_D - g_N)\varphi_i ds, \\ b_i &= \int\int_{\Omega} f\varphi_i dx_1 dx_2, \end{aligned}$$

we can write the system of equations, (2.13) as (we denote nnodes by nn):

$$\begin{aligned} (r_{1,1} + a_{1,1})\xi_1 + \cdots + (r_{1,nn} + a_{1,nn})\xi_{nn} &= rv_1 + b_1, \\ (r_{2,1} + a_{2,1})\xi_1 + \cdots + (r_{2,nn} + a_{2,nn})\xi_{nn} &= rv_2 + b_2, \\ \vdots &= \vdots \\ (r_{nn,1} + a_{nn,1})\xi_1 + \cdots + (r_{nn,nn} + a_{nn,nn})\xi_{nn} &= rv_{nn} + b_{nn}. \end{aligned}$$

In matrix form, this reads,

$$(R + A)\xi = rv + \tilde{b},$$

where

$$R = \begin{bmatrix} r_{1,1} & \cdots & r_{1,nn} \\ \vdots & \ddots & \vdots \\ r_{nn,1} & \cdots & r_{nn,nn} \end{bmatrix}$$

contains the boundary contributions to the system matrix,

$$A = \begin{bmatrix} a_{1,1} & \cdots & a_{1,nn} \\ \vdots & \ddots & \vdots \\ a_{nn,1} & \cdots & a_{nn,nn} \end{bmatrix}$$

is the stiffness matrix,

$$rv = \begin{bmatrix} rv_1 \\ \vdots \\ rv_{nn} \end{bmatrix}$$

contains the boundary contributions to the right-hand side, and

$$\tilde{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_{nn} \end{bmatrix}$$

is the load vector.

*Chapter 3*

# Error Estimation

## 3.1 Introduction

Error Estimation is considered as an important issue in numerical analysis in the sense that it can help us in evaluating approximate the solution or the model itself. Many kinds of errors are presented in the literature, including rounding off error, truncation error, error in data, and uncertainty in the model.
The mathematical theory of estimating discretization error is one of the main and important factor in computational numerical analysis. In fact it can help in assessing the reliability of the result of the computations of the numerical process. The use of measures of error to control time steps in the numerical solution of ordinary differential equations probably represents the first use of a posteriori estimates to control discretization error. The purpose of error estimation is to avoid inaccuracy in the numerical solution, including the errors that come from inaccurate discretization of the solution domain and discretization errors. Also it aims to bound the discretization error $e = u - U$ in a Sobolev space or Lebesgue norm, where $u$ is the exact solution to the variational problem.

$$a(u, v) = f(v), \quad \forall v \in V, \tag{3.1}$$

and $U$ is the approximation solution to the variational problem

$$a(U, v_h) = f(v_h), \forall v_h \in V_h.$$

The error estimate is the difference between approximate solution $U$ and the exact solution $u$, and our task is to test the convergence of the approximate solution to the exact solution as the discretization parameter goes to zero. Normally, the dimensions of the estimating error are similar to that of the solution variable. Also, the kinds of the approximation solution depends on both the discretization parameters and the choice of the exact element space.

Error estimate typically proceeds in two steps, see **[31]**:

$(i)$ Presenting $U$ as a good approximation in the sense that the error $u - U$ satisfies

$$||u - U|| = \min_{v \in V_h} ||u - v|| \tag{3.2}$$

in an appropriate norm, and

$(ii)$ Finding an upper bound for the right-hand side of (3.2). The appropriate norm to use with (3.2) for the model problem (3.1) is the strain energy norm

$$||v||_E = \sqrt{a(v, v)}.$$

The finite element solution might not satisfy (3.2) with other norms. For example, finite element solutions are not optimal in any norm for non-self-adjoint problems, **[31]**. In these cases, (3.2) is replaced by the weaker statement

$$||u - U|| \leq C \min_{v \in V_h} ||u - v||, \quad \text{where } C > 1.$$

Thus, the solution is closed to the best solution but it only differs by a constant from the best possible solution in the space.

Based on finite element approximation, error estimators are usually referred to as explicit error estimators which involve a direct computation of the interior element residuals and the jumps at the element boundaries to find an estimate for the error in the energy norm. In contrast, implicit error estimators require the solution of auxiliary local boundary value problems and involve the solution of the auxiliary boundary value problems whose solution yields an approximation to the actual error, **[32]**. Hence, explicit error estimators in general require less computational effort than implicit schemes. A third class of error estimators is the recovery-based error estimators which make use of the fact that the gradient of the finite element solution is in general discontinuous across the interelement boundaries.

## 3.2 Error estimation of FEM for Poisson Equation

Consider the problem

$$\begin{aligned} -\Delta u &= f, \text{ in } \Omega, \\ u &= 0, \text{ on } \partial\Omega, \end{aligned} \tag{3.3}$$

where $\Omega \subset \mathbb{R}^d, d = 1, 2, 3$. Using Green's theorem,

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega fv \, dx, \quad \forall v \in V_h. \tag{3.4}$$

The variational formulation is : Find $U \in V_h$ such that

$$\int_\Omega \nabla U \cdot \nabla v \, dx = \int_\Omega fv \, dx, \quad \forall v \in V_h. \tag{3.5}$$

For the error $e = u - U$, we have

$$\nabla e = \nabla u - \nabla U = \nabla(u - U).$$

Subtraction of (3.5) from the (3.4) yields the Galarkin Orthogonality

$$\int_\Omega (\nabla u - \nabla U) \cdot \nabla v \, dx = \int_\Omega \nabla e \cdot \nabla v \, dx = 0, \quad \forall v \in V_h. \tag{3.6}$$

On the other hand, we may write

$$||\nabla e||^2 = \int_\Omega \nabla e \cdot \nabla e \, dx = \int_\Omega \nabla e \cdot \nabla u \, dx - \int_\Omega \nabla e \cdot \nabla U \, dx.$$

Now, using the Galarkin Orthogonality (3.6), and since $U \in V_h$, we have

$$\int_\Omega \nabla e \cdot \nabla U \, dx = 0.$$

Employing $\int_\Omega \nabla e \cdot \nabla v \, dx = 0, \ \forall v \in V_h$, to get

$$\begin{aligned} ||\nabla e||^2 &= \int_\Omega \nabla e \cdot \nabla e \, dx \\ &= \int_\Omega \nabla e \cdot \nabla (u - U) \, dx \\ &= \int_\Omega \nabla e \cdot (\nabla u - \nabla U) \, dx \\ &= \int_\Omega \nabla e \cdot \nabla u \, dx - \int_\Omega \nabla e \cdot \nabla U \, dx \\ &= \int_\Omega \nabla e \cdot \nabla u \, dx \\ &= \int_\Omega \nabla e \cdot \nabla u \, dx - \int_\Omega \nabla e \cdot \nabla v \, dx \\ &= \int_\Omega \nabla e \cdot \nabla (u - v) \, dx \\ &\leq ||\nabla e|| \ ||\nabla (u - v)||. \end{aligned}$$

Hence,

$$||\nabla (u - U)|| \leq ||\nabla (u - v)||, \ \ \forall v \in V_h. \tag{3.7}$$

This means that the finite element solution $U \in V_h$ is the best approximation of the solution $u$ among functions in $V_h$, i.e., $U$ is closer to $u$ than any other $v \in V_h$. In this chapter, we shall focus on two types of error estimates for the finite element method, a priori and a posteriori estimates. A priori error estimates are error bounds that use information about the unknown solution $u$ to estimate the error before we compute the approximate solution $U$. They tell us about the order of convergence of a given finite element method, that is, they tell us that the finite element error $||u - U||$ in some norm $|| \cdot ||$ is $O(h^\alpha)$, where $h$ is the maximum mesh size and $\alpha$ is a positive integer. Additionally, the a priori error estimates supply information on convergence rates but are difficult to use for quantitative error information. A posteriori error estimates, which use the computed solution, provide more practical accuracy appraisal, [**31**]. In contrast, a posteriori estimates use the computed solution $U$ in order to give us an estimate of the form $||u - U|| \leq \epsilon$, where $\epsilon$ is a small number.

The main difference between a priori and a posteriori estimates is that a priori error is error bounds given by known information on the solution of the variational problem and the finite element function space. It gives us a reasonable measure of the efficiency of a given method by telling us how fast the error decreases as we decrease the mesh size. But a posteriori estimates are error bounds given by information on the numerical solution obtained on the finite element function space. The a posteriori estimate provides a much better idea of the actual error in a given finite element computation than a priori estimates and it can be used to perform adaptive mesh refinement.

## 3.3 Error estimation in one dimension

### 3.3.1 Dirichlet problem

Assume that a horizontal elastic bar which occupies the interval $I := [0, 1]$, is fixed at the end-points. Let $u(x)$ denote the displacement of the bar at a point $x \in I$, $a(x)$ be the modulus of elasticity, and $f(x)$ a given load function, then one can show that $u$ satisfies the following boundary value problem

$$\begin{aligned} -(a(x)u'(x))' &= f(x), \quad 0 < x < 1, \\ u(0) &= u(1) = 0. \end{aligned} \tag{3.8}$$

Equation (3.8) is of Poisson's type modelling also of the stationary heat flux type. We shall assume that $a(x)$ is piecewise continuous in $(0, 1)$, bounded for $0 \leq x \leq 1$ and $a(x) > 0$ for $0 \leq x \leq 1$.

Let $v(x)$ and its derivative $v'(x), x \in I$, be square integrable functions, that is $v, v' \in L_2(0, 1)$. Define the $L_2$-based Sobolev space:

$$H_0^1(0,1) = \{v(x) : \int_0^1 (v(x)^2 + v'(x)^2)\, dx < \infty, v(0) = v(1) = 0\}$$

The variational formulation (VF) of (3.8) can be obtained by multiplying the equation by a so called test function $v(x) \in H_0^1(0, 1)$ and integrate over $(0, 1)$ to obtain

$$-\int_0^1 (a(x)u'(x))'v(x)dx = \int_0^1 f(x)v(x)\, dx. \tag{3.9}$$

By integration by parts we get

$$-[(a(x)u'(x))'v(x)]_0^1 + \int_0^1 a(x)u'(x)v'(x)dx = \int_0^1 f(x)v(x)\, dx.$$

Now, since $v(0) = v(1) = 0$, the variational formulation for problem (3.8) is as follows: find $u(x) \in H_0^1$ such that

$$\int_0^1 a(x)u'(x)v'(x)dx = \int_0^1 f(x)v(x)\,dx, \;\; \forall v(x) \in H_0^1. \tag{3.10}$$

Continuous Galerkin of degree $1, (cG(1))$ : A finite element formulation for our Dirichlet boundary value problem (3.8) is given by: find $U \in V_h^0$ such that the following discrete variational formulation holds true

$$\int_0^1 a(x)U'(x)'v(x)dx = \int_0^1 f(x)v(x)\,dx, \;\; \forall v \in V_h^0. \tag{3.11}$$

The FEM is a finite dimensional version of the variational formulation, where the test (also trial) functions are in a finite dimensional subspace $V_h^0$, of $H_0^1$, spanned by the hat-functions, $\varphi_j(x), j = 1, \cdots, M$. Thus, if in (3.10) we restrict $v$ to $V_h^0$ (rather than $H_0^1$ ) and subtract the finite element from (3.11), we get the Galerkin orthogonality:

$$\int_0^1 a(x)(u'(x) - U'(x))v(x)dx = 0, \;\; \forall v \in V_h^0. \tag{3.12}$$

### 3.3.2 A mixed Boundary Value Problem

Obviously changing the boundary conditions would require changes in the variational formulation. This can be seen, e.g., in deriving the variational formulation corresponding to the following mixed boundary value problem: find $u$ such that

$$\begin{aligned} -(a(x)u'(x))' &= f(x), \;\; 0 < x < 1, \\ u(0) &= 0, \\ a(1)u'(1) &= g_1. \end{aligned} \tag{3.13}$$

As usual, we multiply the equation by a suitable test function $v(x)$, and integrate over the interval $(0,1)$. Note that, here, the test function should satisfy only one boundary condition: $v(0) = 0$. This is due to the fact that now $u(1)$ is not given, and to get an approximate value of $u$ at $x = 1$, we need to supply a test function (a half-hat-function) at $x = 1$. Therefore, the proper choice for a function space is now

$$\tilde{H_0^1} = \{v(x); \int_0^1 (v(x)^2 + v'(x)^2)\,dx < \infty, \text{ such that } v(0) = 0\}.$$

Let $u \in \tilde{H}_0^1$ multiplying the equation by a test function $v$, such that $v(1) \neq 0$, and integrating over $I = (0,1)$ yield.

$$-\int_0^1 (a(x)u'(x))'v(x)\;dx = \int_0^1 f(x)v(x)\;dx, \;\; \forall v \in \tilde{H}_0^1.$$

Integrating by parts gives

$$-[a(x)u'(x)v(x)]_0^1 + \int_0^1 a(x)u'(x)v'(x)dx = \int_0^1 f(x)v(x)\;dx,$$

and using the boundary data $a(1)u'(1) = g_1$ and $v(0) = 0$ provide

$$-\int_0^1 a(x)u'(x)v'(x)\;dx = \int_0^1 f(x)v(x)\;dx + g_1 v(1), \;\; \forall v \in \tilde{H}_0^1, \tag{3.14}$$

which is the variational formulation of the equation (3.13)

**Error estimates in the energy norm**

We shall study two types of error estimates:

*i*) An a priori error estimate; where a certain norm of the error is estimated by some norm of the exact solution $u$. Here, the error analysis gives information about the size of the error, depending on the (unknown) exact solution $u$, before any computational steps.

*ii*) An a posteriori error estimate; where a certain norm of the error is estimated by some norm of the residual of the approximate solution. The residual is the difference between the left and right hand side of the equation when the exact solution $u(x)$ is replaced by its approximation $U(x)$. Hence, a posteriori error estimates give quantitative information about the size of the error after the approximate solution $U(x)$ has been computed.

Below, we shall prove a qualitative result which shows that the finite element solution is the best approximate solution to the Dirichlet problem in the energy norm.

**Theorem 3.1.** [30] *Let $u(x)$ be the solution to the Dirichlet boundary value problem* (3.8) *and $U(x)$ its finite element approximation given by* (3.11), *then*

$$||u - U||_E \leq ||u - v||_E, \;\; \forall v \in V_h^0.$$

This means that the finite element solution $U \in V_h^0$ is the best approximation of the solution $u$, in the energy norm, by functions in $V_h^0$

*Proof.* We take an arbitrary $\upsilon \in V_h^0$, then using the energy norm

$$\begin{aligned}||u-U||_E^2 &= \int_0^1 a(x)(u'(x)-U'(x))^2\,dx \\ &= \int_0^1 a(x)(u'(x)-U'(x))(u'(x)-\upsilon'(x)+\upsilon'(x)-U'(x))\,dx \\ &= \int_0^1 a(x)(u'(x)-U'(x))(u'(x)-\upsilon'(x))\,dx \\ &+ \int_0^1 a(x)(u'(x)-U'(x))(\upsilon'(x)-U'(x))\,dx\end{aligned}$$

Since $\upsilon - U \in V_h^0 \subset H_0^1$, by Galerkin orthogonality the last integral is zero. Thus,

$$\begin{aligned}||u-U||_E^2 &= \int_0^1 a(x)(u'(x)-U'(x))(u'(x)-\upsilon'(x))\,dx \\ &= \int_0^1 a^{\frac{1}{2}}(x)(u'(x)-U'(x))a^{\frac{1}{2}}(x)(u'(x)-\upsilon'(x))\,dx \\ &\leq \left(\int_0^1 a(x)(u'(x)-U'(x))^2\,dx\right)^{\frac{1}{2}}\left(\int_0^1 a(x)(u'(x)-\upsilon'(x))^2\,dx\right)^{\frac{1}{2}} \\ &= ||u-U||_E\cdot||u-\upsilon||_E \qquad (3.15)\end{aligned}$$

where, in the last estimate, we used Cauchy-Schwarz inequality. Thus

$$||u-U||_E \leq ||u-\upsilon||_E, \quad \forall \upsilon \in V_h^0,$$

and the proof is complete. □

The next step is to show that there exists a function $\upsilon \in V_h^0$ such that $||u-\upsilon||_E$ is not too large. The function that we have in mind is $\phi_h u(x)$ : the piecewise linear interpolant of $u(x)$.

**Theorem 3.2.** **[30]** *[An a priori error estimate]*
*Let $u$ and $U$ be the solutions of the Dirichlet problem* (3.8) *and the finite element problem, respectively. Then there exists an interpolation constant $C_i$, depending only on $a(x)$, such that*

$$||u-\upsilon||_E \leq C_i||hu''||_a.$$

*Proof.* Since $\phi_h u(x) \in V_h^0$, we may take $v = \phi_h u(x)$ in (3.1 and use, e.g., the second estimate in the interpolation

$$\begin{aligned}
||u-U||_E \leq ||u-v|| = ||u-\phi_h u||_E &= ||u' - (\phi_h u)'||_a \\
&= \left( \int_0^1 a(x)(u'(x) - (\phi_h u)'(x)^)2\,dx \right)^{\frac{1}{2}} \\
&\leq \left( \max_{x\in[0,1]} a(x)^{\frac{1}{2}} \right) \cdot ||u' - (\phi_h u)'||_{L_2} \\
&\leq c_i \left( \max_{x\in[0,1]} a(x)^{\frac{1}{2}} \right) ||hu''||_{L_2} \\
&= c_i \left( \max_{x\in[0,1]} a(x)^{\frac{1}{2}} \right) \left( \int_0^1 h(x)^2 u''(x)^2\,dx \right)^{\frac{1}{2}}
\end{aligned}$$

$$||u-U||_E \leq c_i \frac{(\max_{x\in[0,1]} a(x)^{\frac{1}{2}})}{(\min_{x\in[0,1]} a(x)^{\frac{1}{2}})} \left( \int_0^1 a(x) h(x)^2 u''(x)^2\,dx \right)^{\frac{1}{2}}$$

thus

$$C_i = c_i \frac{(\max_{x\in[0,1]} a(x)^{\frac{1}{2}})}{(\min_{x\in[0,1]} a(x)^{\frac{1}{2}})}$$

where $c_i$ is the interpolation constant in the second estimate □

**Remark 3.3.** *If the objective is to divide* $(0,1)$ *into a finite number of subintervals, then one can use the result of Theorem* (3.2)*: to obtain an optimal partition of* $(0,1)$, *where whenever* $a(x)u^{''}(x)^2$ *gets large we compensate by making* $h(x)$ *smaller. This, however, 'requires that the exact solution* $u(x)$ *is known'. Now we shall study a posteriori error analysis, which instead of the unknown solution* $u(x)$, *uses the residual of the computed solution* $U(x)$.

**Theorem 3.4.** **[30]** *(Posteriori error estimate) There is an interpolation constant* $c_i$ *depending only on* $a(x)$ *such that the error in the finite element approximation of the Driichlet boundary value problem* (3.1), *satisfies*

$$||e(x)||_E \leq \left( c_i \int_0^1 \frac{1}{a(x)} h^2(x) R^2(U(x))\,dx \right)^{1/2}$$

*where the residue*

$$R(U(x)) := f + (a(x)U'(x))'$$

*and*

$$e(x) := u(x) - U(x) \in H_0^1.$$

*Proof.* By the definition of the energy norm we have

$$||e(x)||_E^2 = \int_0^1 a(x)(e'(x))^2\,dx = \int_0^1 a(x)(u'(x) - U'(x))e'(x)\,dx$$
$$= \int_0^1 a(x)u'(x)e'(x)\,dx - \int_0^1 a(x)U'(x)e'(x)\,dx. \tag{3.16}$$

Since $e \in H_0^1$, The variational formulation $(VF)_1$ gives that

$$\int_0^1 a(x)u'(x)e'(x)\,dx = \int_0^1 f(x)e(x)\,dx.$$

Hence, we can write

$$||e(x)||_E^2 = \int_0^1 f(x)e(x)\,dx - \int_0^1 a(x)U'(x)e'(x)\,dx$$

Adding and Subtracting the interpolant $\pi_h e(x)$ and its derivative $(\pi_h e)'(x)$ to $e$ and $e'$ in the integrands above yields

$$||e(x)||_E^2 = \int_0^1 f(x)(e(x) - \pi_h e(x))\,dx + \underbrace{\int_0^1 f(x)\pi_h e(x)\,dx}_{(i)}$$
$$- \int_0^1 a(x)U'(x)(e'(x) - (\pi_h e)'(x))\,dx - \underbrace{\int_0^1 a(x)U'(x)(\pi_h e)'(x)\,dx}_{(ii)}$$

Since $U(x)$ is the solution of FEM given by (3.11), and $\pi_h e(x)\ inV_h$ we have that $-(ii) + (i) = 0$. Hence

$$||e(x)||_E^2 = \int_0^1 f(x)(e(x) - \pi_h e(x))\,dx - \int_0^1 a(x)U'(x)(e'(x) - (\pi_h e)'(x))\,dx$$
$$= \int_0^1 f(x)(e(x) - \pi_h e(x))\,dx - \sum_{k=1}^{M+1} \int_{x_{k-1}}^{x_k} a(x)U'(x)(e'(x) - (\pi_h e)'(x))\,dx$$

To continue we integrate by parts in the integrals in the summation above

$$-\int_{x_{k-1}}^{x_k} a(x)U'(x)(e'(x) - (\pi_h e)'(x))\,dx =$$
$$\left[a(x)U'(x)(e(x) - \pi_h e(x))\right]_{x_{k-1}}^{x_k} + \int_{x_{k-1}}^{x_k} (a(x)U'(x))'(e(x) - \pi_h e(x))\,dx.$$

Now, using $e(x_k) = \pi_h e(x_k), k = 0, 1, \cdots, M+1$, where the $x_k s$ are the interpolation nodes, the boundary terms vanish and thus we end up with

$$-\int_{x_{k-1}}^{x_k} a(x)U'(x)(e'(x) - (\pi_h e)'(x))\,dx = \int_{x_{k-1}}^{x_k} (a(x)U'(x))'(e(x) - \pi_h e(x))\,dx.$$

Thus, summing over $k$, we have

$$-\int_0^1 a(x)U'(x)(e'(x) - (\pi_h e)'(x))\,dx = \int_{x_0}^1 (a(x)U'(x))'(e(x) - \pi_h e(x))\,dx.$$

where $(a(x)U'(x))'$ should be interpreted locally on each subinterval $[x_{k-1}, x_k]$. Therefore

$$\begin{aligned} ||e(x)||_E^2 &= \int_0^1 f(x)(e(x) - \pi_h e(x))\,dx + \int_0^1 (a(x)U'(x))'(e(x) - \pi_h e(x))\,dx \\ &= \int_0^1 (f(x) + (a(x)U'(x))')(e(x) - \pi_h e(x))\,dx \end{aligned}$$

Now, let

$$R(U(x)) := f + (a(x)U'(x))',$$

i.e. $R(U(x))$ is the residual error, which is a well-defined except in the set $\{x_k\},\ k = 0, 1, \cdots, M$; where $(a(x_k)U'(x_k))'$ is not defined. Then, using Cauchy-Schwarz inequality we get the following estimate

$$\begin{aligned} ||e(x)||_E^2 &= \int_0^1 R(U(x))(e(x) - \pi_h e(x))\,dx \\ &= \int_0^1 \frac{1}{\sqrt{a(x)}} h(x) R(U(x)) \sqrt{a(x)} \left(\frac{e(x) - \pi_h e(x)}{h(x)}\right)\,dx \\ &= \left(\int_0^1 \frac{1}{a(x)} h^2(x) R^2(U(x))\,dx\right)^{1/2} \left(\int_0^1 a(x) \left(\frac{e(x) - \pi_h e(x)}{h(x)}\right)^2\,dx\right)^{1/2} \end{aligned}$$

Further by definition of the weighted $L_2-$norm we have

$$\left|\left|\frac{e(x) - \pi_h e(x)}{h(x)}\right|\right|_a^2 = \int_0^1 a(x) \left(\frac{e(x) - \pi_h e(x)}{h(x)}\right)^2\,dx \tag{3.17}$$

to estimate (3.17) we can use $||\Pi_h v - v||_{L_p(a,b)} \leq c_i ||hv'||_{L_p(a,b)}$ for $e(x)$ in each subinterval and get

$$\left|\left|\frac{e(x) - \pi_h e(x)}{h(x)}\right|\right|_a \leq C_i \|e'(x)\|_a = C_i \|e(x)\|_E,$$

where $C_i$ as before depends on $a(x)$. Thus

$$||e(x)||_E^2 \leq \left( \int_0^1 \frac{1}{a(x)} h^2(x) R^2(U(x))\, dx \right)^{1/2} C_i \|e(x)\|_E,$$

and the proof is complete. □

## 3.4 Error estimation in two dimensions

**A priori error estimate for poisson equation**

Consider the problem

$$\begin{aligned} -\Delta u &= f, \text{ in } \Omega, \\ u &= 0, \text{ on } \partial\Omega. \end{aligned} \tag{3.18}$$

**Theorem 3.5.** *The finite element approximation $U$ satisfies* (3.18)*. In particular, there is a constant $C_i$ such that*

$$||u - U||_E \leq ||\nabla(u - U)||_E \leq C_i ||hD_u^2|| \tag{3.19}$$

*where $C_i$ is an interpolation constant, and*

$$D^2 u = (u_{xx}^2 + u_{xy}^2 + u_{yy}^2)^{\frac{1}{2}}$$

Now, we will find a priori error estimate for the solution. For a general mesh we have the following a priori error estimate for the solution of the Poisson equation (3.18).

**Theorem 3.6.**

$$||e|| \leq C^2 C_\Omega^2 h^2 ||f||. \tag{3.20}$$

*Proof.* Let $\phi$ be the solution of the dual problem

$$\begin{aligned} -\Delta \phi &= e, \quad \text{in } \Omega, \\ \phi &= 0, \quad \text{on } \partial\Omega. \end{aligned}$$

Then,

$$\begin{aligned} ||e||^2 &= \int_\Omega e \cdot e \, dx \\ &= \int_\Omega e(-\Delta\phi) \, dx \\ &= \int_\Omega \nabla e \cdot \nabla\phi \, dx, \text{by } \ \text{Green's formula} \\ &= \int_\Omega \nabla e \cdot \nabla\phi \, dx - \int_\Omega \nabla e \cdot \nabla v \, dx, \text{by } \ \text{Galarkin Orthogonality} \\ &= \int_\Omega \nabla e \cdot \nabla(\phi - v) \, dx. \end{aligned}$$

So,

$$||e||^2 \leq ||\nabla e|| \, ||\nabla(\phi - v)||, \ \ \forall v \in V_h.$$

Let $v$ be an interpolation of $\phi$ such that

$$||\nabla(\phi - v)|| \leq C||hD^2\phi||,$$

Hence,

$$\begin{aligned} ||e||^2 &\leq ||\nabla e||C||hD^2\phi|| \\ &\leq ||\nabla e||C(\max_\Omega h)||D^2\phi||. \end{aligned} \tag{3.21}$$

To complete the proof, we need the following lemma, see [**30, 33**].

**Lemma 3.7.** *(Regularity Lemma) Assume that $\Omega$ has no re-intrents. We have for $u \in H^2(\Omega)$ with $u = 0$ or $\frac{\partial u}{\partial n} = 0$ on $\partial\Omega$ that,*

$$||D^2u|| \leq C_\Omega||\Delta u||.$$

*Proof.* see ([**30**]) □

Now, applying this lemma to $\phi$,

$$||D^2\phi|| \leq C_\Omega \cdot ||\Delta\phi|| = C_\Omega||e||.$$

Then, (3.21) implies

$$||e||^2 \leq ||\nabla(u - U)||C(\max_\Omega h)C_\Omega||e||.$$

Thus, using

$$||e|| \leq C^2 C_\Omega (\max_\Omega h)||hD^2u||.$$

Which, using the lemma above, for a uniform (constant) $h$, can be written as

$$||e|| \leq C^2 C_\Omega^2 h^2 ||f||.$$

□

### 3.4.1 Dirichlet problem

A posteriori estimates present a necessary tool in the adaptive procedures used in computer simulation and are known to be essential for reliable scientific computing. They are used to control discretization error in numerical solutions of initial or boundary value problems.

**A short history**

The term a posteriori error estimator, was firstly used by Ostrowski [**34**] in 1940. To the authors knowledge, the first use of error estimates for adaptive meshing strategies in significant engineering problems was given in the work of Guerra [**35**] in 1977. The paper of Babuska and Rheinboldt [**36**] published in 1978 is often cited as the first work aimed at developing rigorous global error bounds for finite element approximations of linear elliptic two-point boundary value problems. In the period spanning over two decades since these works, significant advances have been made. A brief history of the subject is given in the book of Ainsworth and Oden [**40**], see also the books and survey articles of Verfurth [**37**], Babuska and Strouboulis [**39**], Oden and Demkowicz [**38**] . It can be argued that until quite recently, the vast majority of the published work on a posteriori error estimation dealt with global estimates of errors in finite element approximations of linear elliptic problems, these estimates generally being in energy-type norms.

The aims of a posteriori error estimation is developing quantitative methods in which the error $e = u - U$ is estimated in post-processing procedures using the solution $U$ as data for the error estimates. A posteriori error estimator is a quantity which bounds or approximates the error and can be computed from the knowledge of numerical solution and input data. The advantage of any a posteriori error estimator is to supply an estimate and ideally bounds for the solution error in a specified norm if the problem data and the finite element solution are available.

### A posteriori error estimate for Poisson equation

To study a posteriori error analysis, where instead of the unknown value of $u(x)$, we use the known value of the approximate solution to estimate the error, [**41**].

**Theorem 3.8.** *Let $u$ be the solution of the Poisson equation* (3.18) *and $U$ is the continuous piecewise linear finite element approximation. Then there is constant $C$, independent of $u$ and $h$, such that*

$$||u - U|| \leq C||h^2 r||, \tag{3.22}$$

*where $r = f + \Delta U$ is the residual.*

*Proof.* Consider the following dual problem

$$\begin{aligned} -\Delta\phi(x) &= e(x), \ x \in \Omega, \\ \phi(x) &= 0, \ x \in \partial\Omega. \end{aligned} \tag{3.23}$$

where it is clear that

$$e(x) = 0, \ \forall x \in \partial\Omega.$$

Using the Green's formula, the $L_2$ norm of the error can be written as

$$||e||^2 = \int_\Omega e^2 \, dx = -\int_\Omega e(\Delta\phi) \, dx = \int_\Omega \nabla e \cdot \nabla\phi \, dx.$$

Thus, by the Galerkin orthogonality and using the boundary condition, we get

$$\begin{aligned} ||e||^2 &= \int_\Omega \nabla e \cdot \nabla\phi \, dx - \int_\Omega \nabla e \cdot \nabla v \, dx \\ &= \int_\Omega \nabla e \cdot \nabla(\phi - v) \, dx. \\ &= \int_\Omega (-\Delta e)(\phi - v) \, dx. \end{aligned}$$

But

$$-\Delta e = -\Delta u + \Delta U = f + \Delta U = r$$

where $r$ is the residual and $v$ is an interpolant of $\phi$, so

$$||e||^2 \leq ||h^2 r|| \; ||h^{-2}(\phi - v)||.$$

Using the inequality

$$||(\phi - v)|| \leq C||h^2 D^2\phi|| \leq CC_\Omega||\Delta\phi||,$$

where $C$ and $C_\Omega$ are constants, we get

$$\begin{aligned} ||e||^2 &\leq CC_\Omega ||h^2 r||\; ||\Delta\phi|| \\ &\leq CC_\Omega ||h^2 r||\; ||e||. \end{aligned}$$

Thus, for this problem, the final a posteriori error estimate is

$$||u - U|| \leq c||h^2 r||.$$

□

### 3.4.2 A mixed boundary condition

Let $\Omega$ be a bounded domain with Lipschitz continuous boundary $\Gamma$. Suppose that $\Gamma$ consists of two measurable parts $\Gamma_D$ and $\Gamma_N$ such that $\Gamma = \Gamma_D \cup \Gamma_D$ where $\Gamma_D$ and $\Gamma_N$ are the Dirichlet and Neumann boundaries, respectively. Consider the mixed boundary value problem: Find a function $u$ such that

$$\begin{aligned} -\Delta u = f, &\quad \text{in } \Omega, \\ u = 0, &\quad \text{on } \Gamma_D, \\ n \cdot \nabla u = g, &\quad \text{on } \Gamma_N, \end{aligned} \tag{3.24}$$

where $n$ is the outward normal to $\Gamma$. We assume that $f \in L_2(\Omega)$ and $g \in L_2(\Gamma_N)$. A variational formulation of this problem is: Find $u \in V$ such that

$$\int_\Omega \nabla u \cdot \nabla v \; dx = \int_\Omega fv \; dx + \int_{\Gamma_N} gv \; ds \;\; \forall v \in V,$$

where the test functions space $V$ is defined as

$$V = v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_D.$$

This solution can be characterized equivalently as the minimizer of the following variational formulation: Find $u \in V$ such that $J(u) = \inf_{v\in V} J(v)$, where

$$J(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 \; dx - \int_\Omega fv \; dx - \int_{\Gamma_N} gv \; ds.$$

To derive the dual variational formulation we employ the relation, [[**27, 28**]],

$$J(u) = \inf_{v\in V} \sup_{y* \in L^2(\Omega,\mathbb{R}^n)} \int_\Omega (\nabla v \cdot y * - \frac{1}{2}|y*|^2 - fv) \; dx - \int_{\Gamma_N} gv \; ds.$$

Define $Q{*}_{f,g} = q{*} \in L^2(\Omega, \mathbb{R}^n)$;

$$\int_\Omega \nabla \cdot q * w \; dx = \int_\Omega -fw \; dx, \int_{\Gamma_N} (q * \cdot n) w \; ds = \int_{\Gamma_N} gw \; ds, \; \forall w \in V,$$

to find $p{*} \in Q_{f,g}{*}$ such that $I * (p{*}) = \sup_{q{*} \in Q_{f,g}{*}} I * (q{*})$, where

$$I * (q{*}) = \int_\Omega (\nabla u \cdot q * -\frac{1}{2}|q * |^2 - fu) \; dx - \int_{\Gamma_N} gu \; ds,$$

is the dual variational functional.
Let

$$\begin{aligned} J(u) &= I * (p{*}), \\ \nabla u &= p{*}, \end{aligned}$$

then we have the following theorem, [**27**].

**Theorem 3.9.** *For all* $v \in V$ *and* $q{*} \in Q{*}_{f,g}$*, we have*
$||\nabla(v-u)||^2 \leq ||\nabla v - q * ||^2, \;\; \forall v \in V, \;\; \forall q{*} \in Q *_{f,g} .$

*Proof.* We will begin as

$$\begin{aligned} J(v) - J(u) &= J(v) - I * (p{*}) \\ &= J(v) - I * (\nabla u) \end{aligned}$$

$$= \int_\Omega (\frac{1}{2}|\nabla v|^2 - fv) \; dx - \int_{\Gamma_N} gv \; ds - (\int_\Omega (\nabla u \cdot \nabla u - \frac{1}{2}|\nabla u|^2 - fu) \; dx - \int_{\Gamma_N} gu \; ds)$$

$$= \int_\Omega (\frac{1}{2}|\nabla(v-u)|^2 + \nabla u \cdot \nabla v - fv - \nabla u \cdot \nabla u + fu) \; dx - \int_{\Gamma_N} (gv - gu) \; ds,$$

but since

$$\begin{aligned} \int_\Omega (\nabla u \cdot \nabla v \; dx &= \int_\Omega fv \; dx + \int_{\Gamma_N} gv \; ds \\ \int_\Omega (\nabla u \cdot \nabla u \; dx &= \int_\Omega fu \; dx + \int_{\Gamma_N} gu \; ds \end{aligned}$$

then we have,

$$J(v) - J(u) = \frac{1}{2}||\nabla(v-u)||^2, \;\; \forall v \in V.$$

Hence, one can derive

$$\begin{aligned}\frac{1}{2}||\nabla(v-u)||^2 &= J(v)-J(u)\\ &= J(v)-I*(p*)\\ &= J(v)-\sup_{q*\in Q*_{f,g}} I*(q*)\\ &= J(v)+\inf_{q*\in Q*_{f,g}} -I*(q*)\\ &= \inf_{q*\in Q*_{f,g}} J(v)-I*(q*).\end{aligned}$$

For the term $J(v)-I*(q*)$ we have

$$J(v)-I*(q*)=\int_\Omega(\frac{1}{2}|\nabla v|^2-fv)\,dx-\int_{\Gamma_N} gv\,ds$$
$$-(\int_\Omega(\nabla u\cdot\nabla u-\frac{1}{2}|\nabla u|^2-fu)\,dx-\int_{\Gamma_N} gu\,ds)$$

$$=\int_\Omega(\frac{1}{2}|\nabla(v-q*)|^2+\nabla q*\cdot\nabla v-fv-\nabla q*\cdot\nabla u+fu)\,dx$$
$$+\int_{\Gamma_N}(gu-gv)\,ds,$$

but

$$\begin{aligned}\int_\Omega q*\cdot\nabla v\,dx &= -\int_\Omega \nabla\cdot q*v\,dx+\int_{\Gamma_N}(q*\cdot n)v\,ds\\ &= \int_\Omega fv\,dx+\int_{\Gamma_N} gv\,ds.\end{aligned}$$

Similarly

$$\int_\Omega q*\cdot\nabla v\,dx=\int_\Omega fu\,dx+\int_{\Gamma_N} gu\,ds.$$

So, we get that

$$J(v)-I*(q*)=\frac{1}{2}||\nabla v-q*||^2,\ \ \forall v\in V,\ \ q*\in Q*_{f,g},$$

and that,

$$||\nabla(v-u)||^2=\inf_{q*\in Q*_{f,g}}||\nabla v-q*||^2.$$

We immediately deduce the estimate

$$||\nabla(v-u)||^2\le||\nabla v-q*||^2,\ \ \forall v\in V,\ \ q*\in Q*_{f,g}\,. \tag{3.25}$$

□

Now we will present a much simplified way of deriving functional type a posteriori estimates using a variant of the Helmholtz decomposition [**25, 26**] for the space $L_2(\Omega, \mathbb{R}^n)$. The Helmholtz decomposition of a vector field is the decomposition of the vector field into two vector fields, one a divergence-free and a curl-free fields. The space$L_2(\Omega, \mathbb{R}^n)$ is used for vector-valued functions with components in $L_2(\Omega)$. Here, we will use the trace theorem, [**24**], that is

$$||u||_{0,\Gamma} \leq C_\Gamma ||u||_{1,\Omega}, \;\; \forall v \in H^1(\Omega) \tag{3.26}$$

where $C_\Gamma$ is positive constants depending only on $\Gamma$, and $||.||_{1,\Omega}$ stands for the standard norm in $H^1(\Omega)$, and the symbol $||.||_{0,\Gamma}$ means the norm is $L_2(\Gamma)$, see, e.g., [**29**].

**Theorem 3.10.** *Let $u \in V$ be the solution to the problem* (3.24) *and $v$ be any function from V. Then, see* [**27, 28**],

$$\begin{aligned} ||\nabla(v-u)||^2 &\leq (1+\beta)||\nabla v - y*||^2 + (1+\frac{1}{\beta}) + \\ &+ (1+\frac{1}{\gamma})C^2_{\Gamma_N}(1+C^2_\Omega||y*\cdot n - g||^2_{L_2(\Gamma_N)}) \\ &+ (1+\frac{1}{\beta})(1+\gamma)C^2_\Omega||div\; y* + f||^2, \end{aligned} \tag{3.27}$$

*where $\beta$ is an arbitrary positive number, $y*$ is any function from $\tilde{H}(\Omega,\ div) = y* \in L_2(\Omega, \mathbb{R}^n) :\ divy* \in L_2(\Omega), y*\cdot n \in L_2(\Gamma_N)$),* [**28**],*$C_\Omega$ is the constant from Ponicare inequality, and $C_{\Gamma_N}$ is the constant in the trace inequality for the domain $\Omega$.*

*Proof.* Consider

$$\begin{aligned} -\Delta u &= f, \text{ in } \Omega, \\ u &= 0, \text{ on } \Gamma_D, \\ n\cdot\nabla u &= g, \text{ on } \Gamma_N, \end{aligned}$$

by (3.25) we have

$$||\nabla(v-u)||^2 \leq ||\nabla v - q*||^2, \;\; \forall v \in V, \;\; \forall q* \in Q*_{f,g}\,.$$

To estimate the right-hand side for any $v \in V$, we take an arbitrary function $y* \in \tilde{H}(\Omega, div)$. Define the auxiliary function $w$ as the solution to the problem

$$\begin{aligned} \Delta w &= divy* + f, \text{ in } \Omega, \\ w &= 0, \text{ on } \Gamma_D, \\ n\cdot\nabla w &= y*\cdot n + g, \text{ on } \Gamma_N. \end{aligned}$$

As $y* \in L_2(\Omega, \mathbb{R}^n)$, we have for $y$ﬃ†the Holmholtz decomposition $y* = q * +\nabla w$, where $q* \in Q*_{f,g}$ and $w \in V$.
Then, using Young's inequality, we obtain

$$||\nabla v - q * ||^2 \leq (1+\beta)||\nabla v - y * ||^2 + (1 + \frac{1}{\beta})||\nabla w||^2, \ \ \forall \beta > 0. \tag{3.28}$$

Since $w \in V$ and $\Delta w \in L_2(\Omega)$, then by Poincare inequality we get

$$\begin{aligned} ||\nabla w||^2 &= \int_{\Gamma_N} \frac{\partial w}{\partial n} w \ ds - \int_\Omega (\Delta w) w \ dx \\ &\leq ||\frac{\partial w}{\partial n}||_{L_2(\Gamma_N)} C_{\Gamma_N} (1 + C_\Omega^2)^{\frac{1}{2}} ||\nabla w|| + C_\Omega ||\Delta w|| \ ||\nabla w||, \end{aligned}$$

that is,

$$||\nabla w|| \leq C_{\Gamma_N} (1 + C_\Omega^2)^{\frac{1}{2}} ||\frac{\partial w}{\partial n}||_{L_2(\Gamma_N)} + C_\Omega ||\Delta w||, \tag{3.29}$$

where $C_\Omega$ is the constant of Poincare inequality, and $C_{\Gamma_N}$ is the constant of the trace inequality. Now by (3.25) we have

$$||\nabla(v - u)||^2 \leq ||\nabla v - q * ||^2, \ \ \forall v \in V, \ \ \forall q* \in Q*_{f,g}$$

Using (**??**) and Young's inequality to get

$$\begin{aligned} ||\nabla(v-u)||^2 &\leq (1+\beta)||\nabla v - y * ||^2 + (1 + \frac{1}{\beta}) \\ &+ (1 + \frac{1}{\gamma}) C_{\Gamma_N}^2 (1 + C_\Omega^2 ||y * \cdot n - g||_{L_2(\Gamma_N)}^2) \\ &+ (1 + \frac{1}{\beta})(1+\gamma) C_\Omega^2 ||div \ y * + f||^2, \ \ \forall v \in V, \ \ \forall y* \in \tilde{H}(\Omega, div), \end{aligned} \tag{3.30}$$

where $\beta$ and $\gamma$ are arbitrary positive numbers come from Young's inequality □

Since $u$ is the exact solution of (3.24), $v$ is any function from $V$, and $y*$ is any function from $\tilde{H}(\Omega, div)$ the estimate (3.30) is an a posteriori error estimate valid for any approximation of the problem (3.24).

*Chapter 4*

# Computations

In this chapter we will test the technique of the FEM to approximate the solutions of differential equations in one and two dimensions

## 4.1 One dimensional examples

**Example 4.1.** *Consider the homogeneous drichlet boundary value problem*

$$\begin{aligned} -u'' &= 12x^2,\ x \in [1,2], \\ u(1) &= u(2) = 0, \end{aligned}$$

*we will test the FEM solution and compare it with the exact solution.*
***Solution:*** *To get the exact solution, we integrate both sides of the differential equation*

$$-u' = 4x^3 + C_1.$$

*Integrate one more both sides to get*

$$-u = x^4 + C_1 x + C_2,\ \ C_1, C_2 \in \mathbb{R},$$

*So,*

$$u(x) = -x^4 + C_1 x + C_2.$$

*Now substitution the boundary conditions,*

$$u(1) = -1 + C_1 + C_2 = 0 \Rightarrow C_1 + C_2 = 1,$$

$$u(2) = -16 + 2C_1 + C_2 = 0 \Rightarrow 2C_1 + C_2 = 16.$$

*The solutions of these two equations provides* $C_1 = 15$ *and* $C_2 = -14$.
*Thus, the exact solution is* $u(x) = -x^4 + 15x - 14$.

Figure (4.1) shows the plot of the exact and approximate solutions at the nodal points. It is clear that refining the mesh provides more accurate result, that is, increasing the number of nodal points decreases the error.

**Example 4.2.** *Consider the one dimensional diffusion-reaction problem*

$$\begin{aligned} -u'' + u &= x^3 - x^2 - 6x + 2, \ x \in [0,1], \\ u'(0) &= 0, \\ u'(1) &= 1, \end{aligned}$$

*compare the finite element solution to the exact solution*
***Solution:*** *To get the exact solution, let* $u_c = e^{rx} \Rightarrow -r^2 + 1 = 0 \Rightarrow r = -1, 1$

$$u_c = C_1 e^x + C_2 e^{-x}, \quad C_1, C_2 \in \mathbb{R}.$$

*Now, the particular solution is*

$$u_p = Ax^3 + Bx^2 + Cx + D \Rightarrow u_p'' = 6Ax + 2B.$$

*Substitute* $u_p$ *and* $u_p''$ *in the differential equation to get*

$$-6Ax - 2B + Ax^3 + Bx^2 + Cx + D = x^3 - x^2 - 6x + 2,$$

*thus* $A = 1, \ B = -1$,
$-6A + C = -6 \Rightarrow C = 0$ *and* $D - 2B = 2 \Rightarrow D = 0$. *Hence,* $u_p = x^3 - x^2$, *and the general solution is*

$$u(x) = C_1 e^x + C_2 e^- x + x^3 - x^2.$$

*Substitute the boundary conditions,*

$$u'(0) = C_1 - C_2 = 0 \Rightarrow C_1 = C_2$$

$$u'(1) = C_1 e - C_2 e^- 1 + 3 - 2 = 1 \Rightarrow C_1 = 0 \text{ which implies } C_2 = 0$$

*Therefore, the exact solution is* $u(x) = x^3 - x^2$

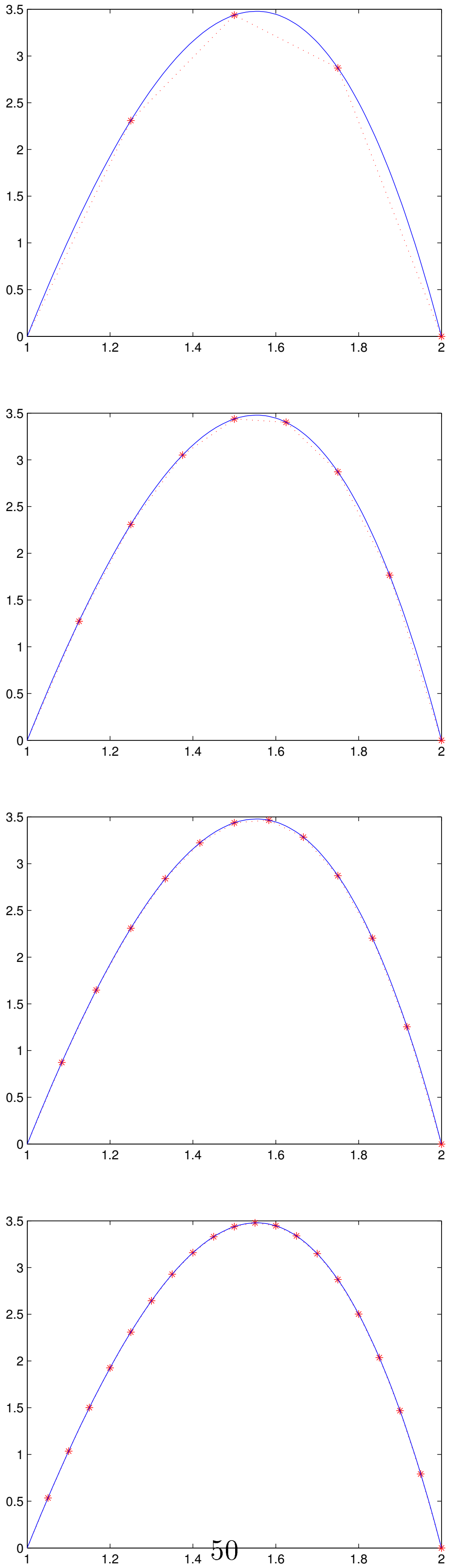


Figure 4.1: Number of subintervals= 4, 8, 12 and 20 respectively.

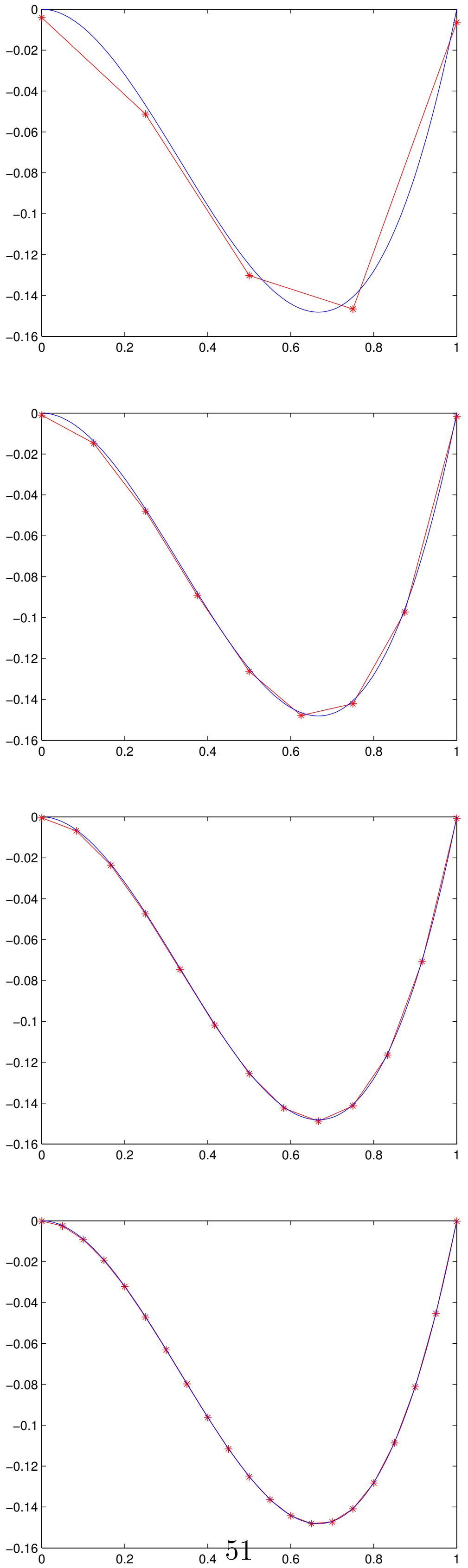


Figure 4.2: Number of subintervals= 4, 8, 12 and 20 respectively.

Figure (4.2) shows the plot of the exact and approximate solutions at the nodal points. The approximate solution is taken for different values of $n$. It is clear that with higher values of subintervals $n$ the solution is more accurate.

This table explains the absolute error in the approximate solution using $4, 8, 12$ and 20 subintervals.

| $x$ | n=4 | n=8 | n=12 | n=20 |
|---|---|---|---|---|
| 0 | 0.004079 | 0.0010095 | 0.0004478 | 0.0001610 |
| 0.25 | 0.004421 | 0.0011007 | 0.0004888 | 0.0001759 |
| 0.5 | 0.005208 | 0.0013020 | 0.0005787 | 0.0002083 |
| 0.75 | 0.005994 | 0.0015034 | 0.0006685 | 0.000240 |
| 1 | 0.006337 | 0.0015946 | 0.0007095 | 0.000255 |

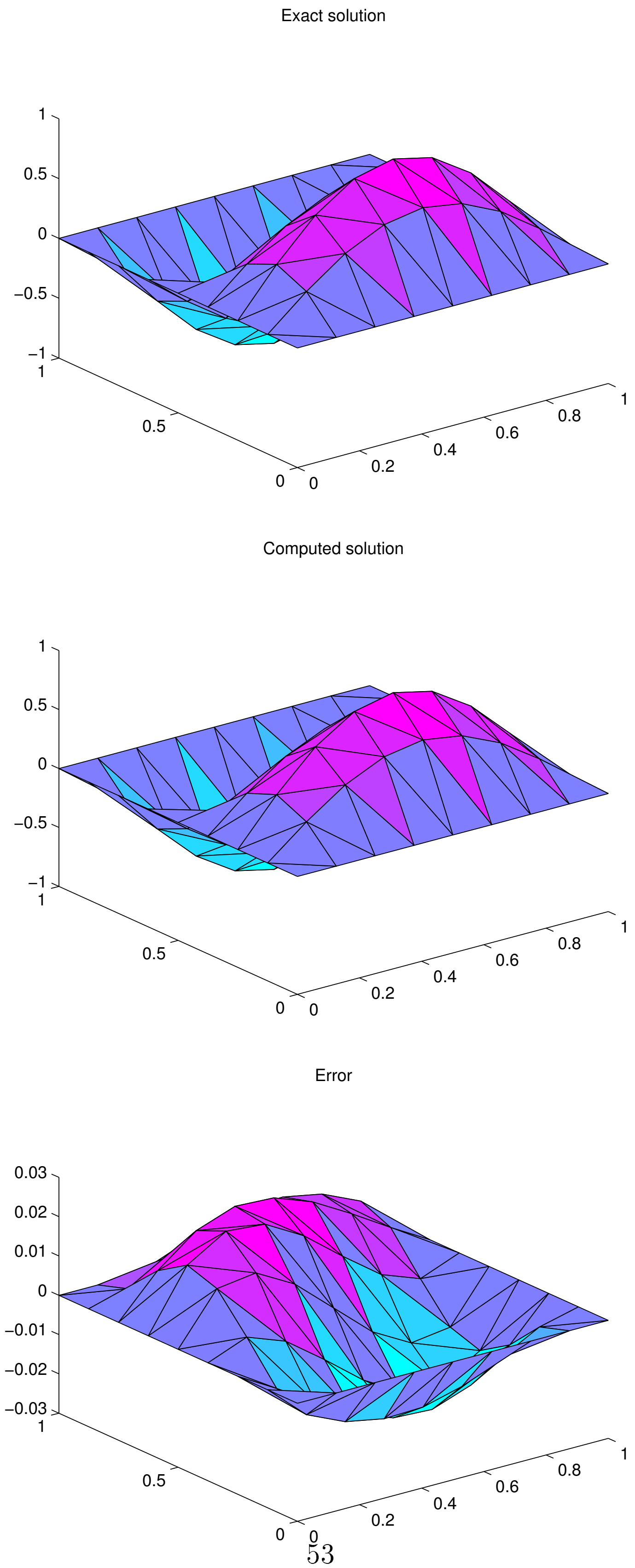
## 4.2 Two dimensional example

Consider the problem

$$\begin{aligned} -\triangle u &= f, \text{ in } \Omega = (0,1) \times (0,1), \\ u &= 0 \text{ on } \partial\Omega. \end{aligned}$$

With right-hand side $f(x) = 5\pi^2 \sin(\pi x_1) \sin(2\pi x_2)$. where exact solution $u(x) = \sin(\pi x_1) \sin(2\pi x_2)$

The error is plotted in Figure (4.3), where the maximum norm it is equal to 0.0216. In Figure (4.4) are refine the mesh and obtain the approximate solution. One can note that the maximum norm of the error is 0.0055. It is clear from Figures (4.3) and (4.4) that refining the mesh gives better approximation. This is reasonable because the more mesh there are, the less error. Note that the error is decreased by the factor $4 = 2^2$ when we decrease the mesh size with a factor 2.

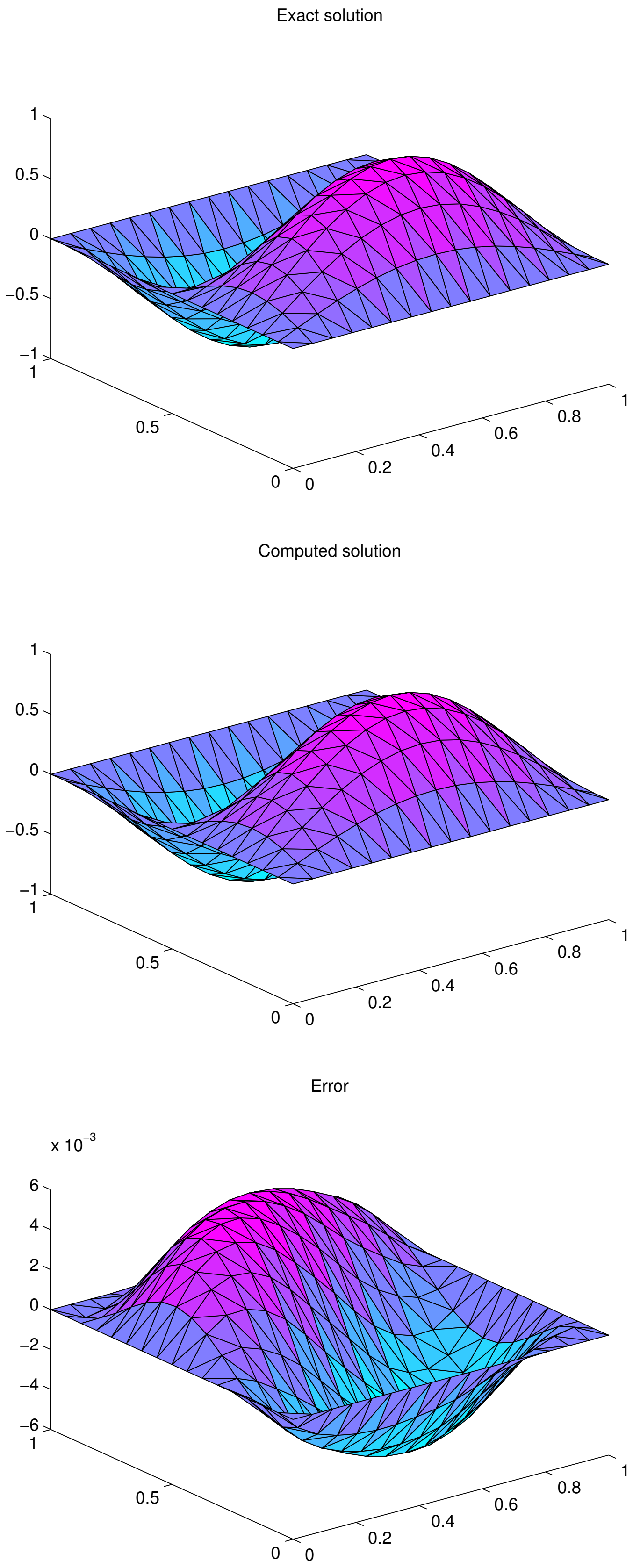


Exact solution

Computed solution

Error

Figure 4.3: Exact solution, approximate solution, and error.



Exact solution

Computed solution

Error

Figure 4.4: Exact solution, approximate solution, and error.

# Bibliography

[1] Bathe, K.-J. (2014). Finite Element Procedures.

[2] Braess, D. (2007a). Finite elements. Theory, fast solvers and applications in elasticity theory. (Finite Elemente. Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie.) 4th revised and extended ed.

[3] Braess, D. (2013). Finite elemente: Theorie, schnelle löser und anwendungen in der elastizitätstheorie.

[4] Brenner, S. and Scott, R. (2008). The Mathematical Theory of Finite Element Methods.

[5] Burden, R. L., Faires, J. D., and Burden, A. M. (2015). Numerical Analysis.

[6] Cao, Y., Nie, S., and Wu, Z. (2019). Numerical simulation of parachute inflation:A methodological review. Proceedings of the Institution of Mechanical Engineers,Part G: Journal of Aerospace Engineering, 233(2):736-766.

[7] Ciarlet, P. G. (2002). Finite element method for elliptic problems.

[8] Eriksson, K. (1996). Computational differential equations.

[9] Evans, L. C. (2010). Partial Differential Equations: Second Edition.

[10] Gaeta, G. and RodrÃguez, M. A. (2017). Lectures on Hyperhamiltonian Dynamics and Physical Applications.

[11] Izadi, M. (2007). Streamline diffusion method for treating coupling equations of hyperbolic scalar conservation laws. Mathematical and Computer Modelling,45(1):201-214.

[12] Kuo, H.-J. and Trudinger, N. S. (1992). Discrete methods for fully nonlinear elliptic equations. 29(1):123-135.

[13] Langtangen, H. P. and Mardal, K.-A. (2019). Introduction to Numerical Methods for Variational Problems.

[14] Larson, M. G. and Bengzon, F. (2013). The Finite Element Method: Theory,Implementation, and Applications.

[15] Quarteroni, A. (2014). Numerical models for differential problems.

[16] Quarteroni, A. M. and Valli, A. (2008). Numerical Approximation of Partial Differential Equations.

[17] Renardy, M. and Rogers, R. C. (2004). An introduction to partial differential equations.

[18] Saad, Y. (2003). Iterative Methods for Sparse Linear Systems.

[19] Thomas, J. W. (1998). Numerical partial differential equations: Finite difference methods.

[20] Wait, A. R. M. (1997). Finite Element Method in Partial Differential Equations by A. Richard Mitchell.

[21] Wu, J. (1996). Theory and applications of partial functional differential equations.

[22] Zeidler, E. (2007). Quantum Field Theory I: Basics in Mathematics and Physics:A Bridge between Mathematicians and Physicists. Google-Books-ID: XYtnGl9enNgC.

[23] Zhang, Z. and Yan, N. (2001). Recovery type a posteriori error estimates in finite element methods. 8(2):235.

[24] S. Korotov, A posteriori error estimation for linear elliptic problems with mixed boundary conditions, Helsinki University of Technology, Institute of Mathematics, Research Reports A495 (2006).

[25] S.I Repin, A posteriori error estimation for nonlinear variational problems by duality theory, Zapiski Nauchnih Seminarov, V.A. Stekov Mathematical Institute (POMI), pp. 201-214, 243(1997).

[26] S.I. Repin, A posteriori error estimation for variational problems with uniformly convex functionals, Math. Comp. 69,pp. 481-500, , 230 (2000).

[27] S. Repin, S. Sauter and A. Smolianski, A posteriori error estimation for the Dirichlet problem with account of the error in the approximation of boundary conditions, Computing 70, pp.205-233, (2003).

[28] S. Repin, S. Sauter, A. Smolianski, A posteriori error estimation for the Poisson equation with mixed Dirichlet/Neumann boundary conditions, Journal of Computational and Applied Mathematics, pp. 601-612 ,164 (2004).

[29] J. Necas, Les methodes directes en theorie des equations elliptiques, Academia, Praha, and Masson et Cie, Editeurs, Paris, (1967).

[30] Asadzadeh, An Introduction to the Finite Element Method (FEM) for Differential Equations, (2012).

[31] J. E. Flaherty, Finite element analysis lecture notes, spring, CSCI, math 6860, Troy, New York 12180, (2000).

[32] T.Gratsch, K. Bathe, A posteriori error estimation techniques in practical finite element analysis, Computers and Structures, pp. 235-265, 83 (2005).

[33] K. Eriksson, D. Estep, P. Hansbo and C. Johnson, Computational Differential Equations,(2009).

[34] A. Ostrowski, Recherches sur la methode de Graeffe et les zeros des polynomes et des series des Laurent. Acta Math., pp. 99-257, 72 (1940).

[35] F. M. Guerra, Finite element analysis for the adaptive method of rezoning, PhD thesis, The University of Texas at Austin, (1977).

[36] I. Babuska and W.C. Rheinboldt, A Posteriori Error Estimate for the Finite Element Method, SIAM J. Numer. Anal., pp. 1597-1615, 15 (1978).

[37] R. Verfurth, A Review of A Posteriori Error Estimation and Adaptive Mesh-refinement Techniques, Wiley-Teubner, Stuttgart, (1996).

[38] J. T. Oden and L. Demkowicz, A survey of adaptive finite element methods in computational mechanics, In A. K. Noor and J. T. Oden, editors, State-of-the-Art Surveys on Computational Mechanics. The American Society of Mechanical Engineers, New York, (1989).

[39] I. Babuska and T. Strouboulis. The Finite Element Method and Its Reliability, Oxford University Press Inc., New York, (2001).

[40] M. Ainsworth and J.T. Oden, A Posteriori Error Estimation in Finite Element Analysis, John Wiley and Sons Interscience, New York, 2000.

[41] T.Gratsch, K. Bathe, A posteriori error estimation techniques in practical finite element analysis, Computers and Structures, pp. 235-265, 83 (2005).