**UNIVERSITY OF BOTSWANA**
**DEPARTMENT OF MATHEMATICS**

# LECTURE NOTES ON DISTRIBUTIONS VARIATIONAL CALCULUS POLYNOMIAL SPLINES AND FINITE ELEMENT METHOD WITH MATHEMATICA

**TADEUSZ STYŠ**

**Gaborone, May 2005**

# PREFACE

This text is intended for science and engineering students pursuing a post graduate course on solution of ordinary and partial differential equations. It presents theoretical background and principles for variational methods and finite element methods with application of polynomial splines. As a pri-requisite material, it requires basic knowledge in mathematical analysis, differential equations, numerical analysis and computing. The lecture notes cover the following related topics:

- Distributions
- Variational Calculus
- Polynomial splines
- Finite element methods

There is extensive literature published on variational methods and finite element methods, (cf. [1,2,3,4,5,6,7,8,9]) In this text, we present a compact lecture notes on theory and application of the methods based on variational principles.
The distributions, weak solutions of differential equations and spline functions are naturally associated with calculus of variations and finite element methods. We present these areas of Mathematics in the above listed four chapters. Each chapter ends with a number of exercises. It is taken for granted that the reader will have access to computer facilities to aid in the solving of some of the exercises.
Most of the material of this book has its origin based on lecture courses given to advanced undergraduate and postgraduate students.
There is also a hope that users will find a number of interesting algorithms, many variants of polynomial splines and useful models of finite element method. Basic theory of the methods has been amplified and designed for those readers who will wish to study effective techniques of solution of differential equations. The text contains lengthy proofs and formulae. They are not necessary to understand the methods and to use them practically, but they lead to better theoretical knowledge.

Tadeusz STYŠ

# Contents

## List of Mathematica Modulae

# Chapter 1

# Distributions

## 1.1 Schwarz's Definition of Distributions

Let $\Omega \subseteq R^n$ be a domain and let $C_0^\infty(\Omega)$ be the set of all infinitely differentiable functions which have support in $\Omega$, i.e., $\psi \in C_0^\infty(\Omega)$ if and only if

1. (a) $\psi$ has all derivatives in $\Omega$,

   (b) *supp* $\psi(x) \subset \Omega$,

where

$$supp\ \psi(x) = \overline{\{x = (x_1, x_2, ..., x_n) \in \Omega\ :\ \ \psi(x) \neq 0\}}.$$

We note that every function $\psi \in C_0^\infty(\Omega)$ vanishes at the boundary $\partial\Omega$ of the domain $\Omega$ together with all its derivatives.

**Example 1.1** *Let us consider the following function:*

$$\psi_a(x) = \begin{cases} \exp(-\dfrac{a^2}{a^2 - x^2}) & if \quad |x| < a, \quad a > 0, \\ 0 & if \quad |x| \geq a, \end{cases}$$

*This function has all its derivatives at each $x \in \Omega = R$ and*

$$supp\ \psi_a(x) = [-a, a] \subset R$$

*Therefore, $\psi_a \in C_0^\infty(\Omega)$.*

Below, we give Schwarz's definition of a distribution.

**Definition 1.1** *A linear functional $F(\psi)$, $\psi \in C_0^\infty(\Omega)$ is said to be a distribution in the domain $\Omega$ if and only if for every compact domain $S \subset \Omega$ there exist constants $m$ and $K$ such that*

$$|F(\psi)| \leq K \sum_{|\alpha| \leq m} \sup_{x \in S} |D^\alpha \psi(x)|$$

*for any* $\psi \in C_0^\infty(S)$*, where the multi subscript* $\alpha = (\alpha_1, \alpha_2, ..., \alpha_n)$*, and the derivative*

$$D^\alpha u(x) = \frac{\partial^{|\alpha|} u(x)}{\partial x_1^{\alpha_1} \partial x_2^{\alpha_2} \cdots \partial x_n^{\alpha_n}}, \qquad |\alpha| = \alpha_1 + \alpha_2 + \cdots + \alpha_n.$$

## 1.2 Dirac $\delta$ and Heaviside Distributions.

**Dirac's** $\delta$. Dirac's $\delta$ was introduced by Dirac in a model equation with a source of energy concentrated at a point. Then, $\delta$ was defined by the following conditions:

1. $\delta(x) = 0 \quad \text{for} \quad x \neq 0,$
2. $\int_{-\infty}^{\infty} \delta(x) dx = 1.$

The quantity $\delta$ determined by conditions 1 and 2 is not a function, but it is an original distribution. In the theory of distributions, Dirac's $\delta$ is considered as a distribution defined by the following formula:

$$\delta(\psi) = \psi(0) \qquad \text{for any} \quad \psi \in C_0^\infty(R). \tag{1.1}$$

Let us check that $\delta$ given by formula (1.1) is a distribution in the sense of Schwarz's definition (1.1). Indeed, $\delta$ is a linear functional, since

$$\delta(\lambda_1\psi_1 + \lambda_2\psi_2) = \lambda_1\psi_1(0) + \lambda_2\psi_2(0) = \lambda_1\delta(\psi_1) + \lambda_2\delta(\psi_2).$$

for every numbers $\lambda_1, \lambda_2$ and $\psi_1, \psi_2 \in C_0^\infty(\Omega)$.
Also, $\delta$ satisfies the inequality

$$|\delta(\psi)| = |\psi(0)| \leq \sup_{x\in S} |\psi(x)|,$$

for the constants $m = 0,\ K = 1$ and any $\psi \in C_0^\infty(R)$.
Let us note that $\delta - \textit{distribution}$ is a limit of the sequence of the following functions:

$$\delta_n(x) = \begin{cases} 0 \quad if \quad x < -\dfrac{1}{n} \quad or \quad x > \dfrac{1}{n}, \\ \dfrac{n}{2} \quad if \quad -\dfrac{1}{n} \leq x \leq \dfrac{1}{n}, \end{cases}$$

for n=1,2,...;

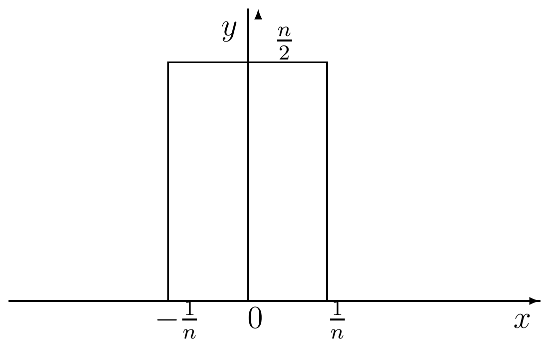



Fig. 7.1. $\delta_n(x)$

Indeed

$$\int_{-\infty}^{\infty} \delta_n(x)dx = 1 \qquad \text{for every} \quad n = 1, 2, ...;$$

and

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \delta_n(x)dx = 1.$$

Hence

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} \delta_n(x)\psi(x)dx = \frac{n}{2} \lim_{n\to\infty} \int_{-\frac{1}{n}}^{\frac{1}{n}} \psi(x)dx = \psi(0),$$

for any $\psi \in C_0^\infty(R)$.

**Heaviside's distribution.** Let us consider the following functional:

$$F_H(\psi) = \int_0^\infty \psi(x)dx,$$

for any $\psi \in C_0^\infty(R)$.

The functional $F_H(\psi)$ is called Heaviside's distribution. Let us note that $F_H(\psi)$ satisfies definition (1.1). Namely, $F_H(\psi)$ is a linear functional, since

$$F_H(\lambda_1\psi_1 + \lambda_2\psi_2) = \int_0^\infty [\lambda_1\psi_1(x) + \lambda_2\psi_2(x)]dx =$$
$$= \lambda_1 \int_0^\infty \psi_1(x)dx + \lambda_2 \int_0^\infty \psi_2(x)dx = \lambda_1 F_H(\psi_1) + \lambda_2 F_H(\psi_2).$$

for arbitrary numbers $\lambda_1, \lambda_2$ and $\psi_1, \psi_2 \in C_0^\infty(R)$.

Also, we have

$$|F_H(\psi)| = |\int_0^\infty \psi(x)dx| \le K \max_{x\in S} |\psi(x)|, \qquad \psi \in C_0^\infty(R).$$

where $K$ is a measure of the support $supp\, \psi(x)$.

## 1.3 Generalized Functions

Generalized functions are distributions which are determined by locally integrable functions.
Let f be a locally integrable function in the domain $\Omega$, i.e., there exists the integral

$$\int_S f(x)dx < \infty$$

for any compact domain $S \subset \Omega$.
Then, we consider the linear functional

$$F_f(\psi) = \int_\Omega f(x)\psi(x)dx$$

for any $\psi \in C_0^\infty(\Omega)$.
The functional $F_f(\psi)$ is a distribution in the sense of definition (1.1) determined by the locally integrable function $f$. Indeed, $F_f(\psi)$ is a bounded linear functional, since

$$|F_f(\psi)| = |\int_S f(x)\psi(x)dx| \leq K \sup_{x\in S} |\psi(x)|,$$

for any $\psi \in C_0^\infty(S)$ and a compact domain $S \subset \Omega$, where

$$K = \int_S |f(x)|dx.$$

Thus, every locally integrable function $f$ determines a distribution $F_f(\psi)$, $\psi \in C_0^\infty(\Omega)$.
Obviously, the locally integrable functions $f$ and $g$ determine the same distribution, i.e., $F_f(\psi) = F_g(\psi)$ for any $\psi \in C_0^\infty(\Omega)$, if $f(x) = g(x)$ almost everywhere in $\Omega$.
[1] However, there exist certain distributions which are not generalized functions. For instance, $\delta - \textit{distribution}$ is not a generalized function, since it is not determined by any locally integrable function. Indeed, if $\delta$ is a generalized function then

$$\delta(\psi) = \int_{-\infty}^{\infty} f(x)\psi(x)dx \tag{1.2}$$

for any $\psi \in C_0^\infty(R)$ and certain locally integrable function $f$ in $R$.
On the other hand $\delta(\psi) = \psi(0)$. Therefore

$$\int_{-\infty}^{\infty} f(x)\psi(x)dx = \psi(0) \tag{1.3}$$

[1] $f(x) = g(x)$ almost everywhere in $\Omega$ if this equality holds for all $x \in \Omega - \Omega_0$, where $\Omega_0$ has a measure equal to zero.

for any $\psi \in C_0^\infty(R)$. Let

$$\psi(x) = \begin{cases} \exp(-\dfrac{a^2}{a^2 - x^2}) & if \quad |x| \le a, \quad a > 0, \\ 0 & if \quad |x| > a. \end{cases}$$

Then, by (1.3)

$$\int_{-\infty}^{\infty} f(x)\psi(x)dx = \int_{-a}^{a} f(x)\exp(-\frac{a^2}{a^2 - x^2})dx = \exp(-1). \tag{1.4}$$

But, we can choose such $a$ that

$$|\int_{-a}^{a} f(x)\exp(-\frac{a^2}{a^2 - x^2})dx| < \exp(-1).$$

The last inequality contradicts equality (1.4). Therefore, there is not a locally integrable function on $R$ for which the equality (1.2) holds.

**Derivatives of distributions.** Let $F(\psi)$ be a distribution given in the domain $\Omega$. Then, the first derivative of $F(\psi)$ with respect to $x_j$, $j = 1, 2, ..., n$; is defined by the formula

$$\frac{\partial F(\psi)}{\partial x_j} = -F(\frac{\partial \psi}{\partial x_j}), \qquad j = 1, 2, ..., n, \tag{1.5}$$

for any $\psi \in C_0^\infty(\Omega)$.

If the distribution $F(\psi)$ is a generalized function determined by a locally integrable function $f$ then, by (1.5), its derivative

$$\frac{\partial F_f(\psi)}{\partial x_j} = -\int_\Omega f(x)\frac{\partial \psi}{\partial x_j}dx, \qquad j = 1, 2, ..., n.$$

We can thus say that a derivative of a distribution is also a distribution.

Higher order derivatives of a distribution are determined by the following formula:

$$\frac{\partial^k F(\psi)}{\partial x_j^k} = (-1)^k F(\frac{\partial^k \psi}{\partial x_j^k}), \qquad j = 1, 2, ..., n, \tag{1.6}$$

for any $\psi \in C_0^\infty(\Omega)$ and integer $k$.

Hence, the derivative of a generalized function is:

$$\frac{\partial^k F_f(\psi)}{\partial x_j^k} = (-1)^k \int_\Omega f(x)\frac{\partial^k \psi}{\partial x_j^k}dx, \qquad j = 1, 2, ..., n. \tag{1.7}$$

for any $\psi \in C_0^\infty(\Omega)$ and an integer $k$.

**Example 1.2** *Find all derivatives of Heaviside's distribution*

$$F_H(\psi) = \int_0^\infty \psi(x)dx$$

.

Since Heaviside's distribution is a generalized function determined by the following locally integrable function:

$$H(x) = \begin{cases} 1 & if \quad x > 0, \\ 0 & if \quad x \leq 0, \end{cases}$$

by the formula (1.6), its first derivative

$$\frac{dF_H(\psi)}{dx} = -\int_0^\infty \frac{d\psi}{dx}dx = -\psi(\infty) + \psi(0) = \psi(0) = \delta(\psi).$$

Thus, the first derivative of Heaviside's distribution is $\delta$ distribution.
In order to find higher order derivatives of $F_H(\psi)$, we apply formula (1.7).
Then, we have

$$\frac{d^k F_H(\psi)}{dx^k} = (-1)^k \int_0^\infty \frac{d^k\psi}{dx^k}dx = (-1)^k[\frac{d^{k-1}\psi(\infty)}{dx^{k-1}} - \frac{d^{k-1}\psi(0)}{dx^{k-1}}] = (-1)^{k-1}\frac{d^{k-1}\psi(0)}{dx^{k-1}},$$

for $k \geq 1$.
Let us note that $H(x)$ is a discontinuous function at $x = 0$. However, $H(x)$ determines Heaviside's distribution $F_H(\psi)$ which has all derivatives.

**Example 1.3** *Let us consider the following distributions determined by the trigonometric functions:*

1. $F_{sin}(\psi) = \int_{-\infty}^\infty \psi(x) \sin\, x\, dx,$

2. $F_{cos}(\psi) = \int_{-\infty}^\infty \psi(x) \cos\, x\, dx,$

By formula (1.7), we obtain

$$\frac{F_{sin}(\psi)}{dx} = -\int_{-\infty}^\infty \frac{d\psi}{dx} \sin x\; dx = -[\psi(x)\sin x|_{-\infty}^\infty + \int_{-\infty}^\infty \psi(x) cos\; x\; dx = F_{cos}(\psi),$$

$$\frac{F_{cos}(\psi)}{dx} = -\int_{-\infty}^\infty \frac{d\psi}{dx} \cos x\; dx = -[\psi(x)\cos x|_{-\infty}^\infty - \int_{-\infty}^\infty \psi(x)\sin\; x\; dx = -F_{sin}(\psi),$$

So that

$$\frac{dF_{sin}(\psi)}{dx} = F_{cos}(\psi), \qquad \frac{dF_{cos}(\psi)}{dx} = -F_{sin}(\psi).$$

Let us now consider derivatives of the generalized function $F_u(\psi)$. If there exists a locally integrable function $v$ in the domain $\Omega$ such that

$$\int_\Omega u \frac{\partial^{|\alpha|}\psi}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n}} dx = (-1)^{|\alpha|}\int_\Omega v(x)\psi(x)dx$$

for any $\psi \in C_0^\infty(\Omega)$, then the generalized function

$$F_v(\psi) = (-1)^{|\alpha|}\int_\Omega v(x)\psi(x)dx$$

is the generalized derivative of $u$ of order $|\alpha|$.

**Example 1.4** *The function* $u = \exp(x_1 + x_2)$ *determines the distribution*

$$F_u(\psi) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(x_1 + x_2)\psi(x_1, x_2)dx_1dx_2, \qquad \psi \in C_0^\infty(R^2).$$

*Because*

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(x_1 + x_2)\frac{\partial^2\psi}{\partial x_1\partial x_2}dx_1dx_2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp(x_1 + x_2)\psi(x_1, x_2)dx_1dx_2.$$

*therefore, the generalized derivative of* $u = u = \exp(x_1 + x_2)$ *is the same function* $u$.

## 1.4 Weak Solutions

Let us consider equation (2.1)

$$\frac{du}{dx} = f(x), \qquad x \in (0, 1), \tag{1.8}$$

where

$$f(x) = \begin{cases} 1 \quad if \quad 0 \le x \le \dfrac{1}{2}, \\ 0 \quad if \quad \dfrac{1}{2} < x \le 1, \end{cases}$$

As we know, $f(x)$ determines the distribution

$$F_f(\psi) = \int_0^1 \psi(x)dx.$$

Thus, we can write equation (1.8) in terms of distributions

$$\frac{dF_u(\psi)}{dx} = F_f(\psi), \qquad \psi \in C_0^\infty(0, 1).$$

or in the equivalent form

$$-\int_0^1 u\frac{d\psi}{dx}dx = \int_0^1 f(x)\psi(x)dx, \qquad \psi \in C_0^\infty(0,1). \tag{1.9}$$

Let us observe that the function

$$u(x) = \begin{cases} x \quad if \quad 0 \le x \le \frac{1}{2}, \\ \frac{1}{2} \quad if \quad \frac{1}{2} < x \le 1, \end{cases}$$

satisfies the variational equation (1.9). Therefore, the distribution

$$F_u(\psi) = \int_0^1 u(x)\psi(x)dx, \qquad \psi \in C_0^\infty(0,1),$$

is called a weak solution of the differential equation (1.8). This solution is not differentiable at $x = \frac{1}{2}$. So, $u(x)$ cannot be a regular solution of equation (1.8). The function $u(x)$ is the most regular representation of the distribution $F_u(\psi)$.

We shall now introduce the idea of weak solutions of a linear differential equation. A linear differential equation can be written in the following form:

$$Lu \equiv \sum_{|\alpha|\le k} A_\alpha(x)\frac{\partial^{|\alpha|}u}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n}} = f(x), \qquad x \in \Omega, \tag{1.10}$$

where the coefficients $A_\alpha \in C^k(\Omega)$ and $f$ is a locally integrable function in $\Omega$. Then, the Lagrange's conjugate differential equation to (1.10) is:

$$L^*v \equiv \sum_{|\alpha|\le k} (-1)^{|\alpha|}\frac{\partial^{|\alpha|}(A_\alpha v)}{\partial x_1^{\alpha_1}\partial x_2^{\alpha_2}\cdots\partial x_n^{\alpha_n}} = f(x), \qquad x \in \Omega.$$

Let us write the variational form of the differential equation (1.10) as follows:

$$\int_\Omega u(x)L^*\psi(x)dx = \int_\Omega f(x)\psi(x)dx, \tag{1.11}$$

for any $\psi \in C_0^\infty(\Omega)$.

Then, the distribution

$$F_u(\psi) = \int_\Omega u(x)\psi(x)dx, \qquad \psi \in C_0^\infty(\Omega)$$

determined by (1.11) is a weak solution of the differential equation (1.10).

**Example 1.5** *Let us consider the Poisson equation*

$$\Delta u \equiv \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = \begin{cases} 0 \quad if \quad x_2 \ne 1 - x_1, \\ 1 \quad if \quad x_2 = 1 - x_1, \end{cases} \qquad x = (x_1, x_2) \in \Omega, \tag{1.12}$$

*with the homogeneous boundary condition, i.e.,* $u(x,y) = 0$ *for* $x \in \partial\Omega$*, where* $\Omega = \{x = (x_1, x_2) \ : \ 0 < x_1, x_2 < 1\}$.

The Poisson equation does not have a differentiable solution. However, it possesses a weak solution. Namely, the Lagrange's conjugate equation to (1.12) is the same Poisson's equation. Therefore its variational form is:

$$\int_\Omega u(x_1,x_2)\Delta\psi(x_1,x_2)dx_1dx_2 = \int_\Omega \delta(x_1+x_2-1)\psi(x_1,x_2)dx_1dx_2$$

for any $\psi \in C_0^\infty(\Omega)$, where

$$f(x_1,x_2) = \begin{cases} 0 & if \quad x_2 \neq 1-x_1, \\ 1 & if \quad x_2 = 1-x_1. \end{cases}$$

Then, the continuous function

$$u(x_1,x_2) = \begin{cases} x_1x_2 & if \quad x_2 \leq 1-x_1, \\ (1-x_1)(1-x_2) & if \quad x_2 > 1-x_1, \end{cases}$$

is the most regular representative of the weak solution

$$F_u(\psi) = \int_\Omega u(x_1,x_2)\psi(x_1,x_2)dx_1x_2$$

of Poisson's equation (1.12).

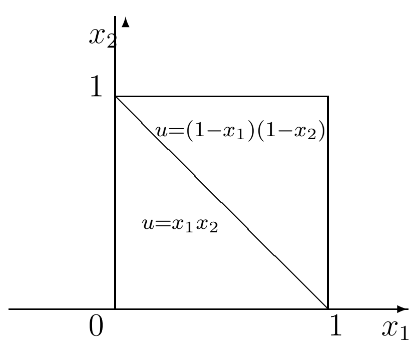



Fig. 7.2. $u(x_1,x_2)$

**Sobolev's space $W_m^l(\Omega)$.** The Sobolev's space $W_m^l(\Omega)$ is naturally associated with both regular and weak solutions of differential equations. This space can be introduced in the following two ways:

1.
$$W_m^l(\Omega) = \{u \ : \ D^{|\alpha|}u \in L_m(\Omega), \ \ |\alpha| \leq l, \}$$

where

$$L_m(\Omega) = \{u \ : \ [\int_\Omega |u(x)|^m d\Omega]^{\frac{1}{m}} < +\infty, \quad m \geq 1\}.$$

2. The space $W_m^l(\Omega)$ is the closure of the set $C_0^\infty(\Omega)$ in the norm

$$||u|| = [\int_\Omega \sum_{|\alpha|\leq l} |D^{|\alpha|}u|^m d\Omega]^{\frac{1}{m}}, \qquad m \geq 1.$$

For a sufficiently regular domain $\Omega$, conditions 1 and 2 are equivalent. Later on, we shall also use the subspace $W_m^{0l}(\Omega)$, $m \geq 2,\ l \geq 1$ of the space $W_m^l(\Omega)$. The subspace $W_m^{0l}(\Omega)$ can be obtained as a closure of the set $C_0^\infty(\Omega)$ in the norm of the Sobolev's space $W_m^l(\Omega)$.
Let us note that every function $u \in W_m^{0l}(\Omega)$ vanishes at the boundary $\partial\Omega$ of the domain $\Omega$ together with all its derivatives up to the order $l-1$. For instance, if $u \in W_2^{01}(R)$, then $u$ vanishes at $\pm\infty$. In fact, elements of the Sobolev's space $W_m^l(\Omega)$ are generalized functions. where the integers $m$ and $l$ determine regularity of $u$. For example, if a generalized function $u \in W_2^{01}$, then there exists a continuous function $u$ in $C(\Omega)$ which vanishes at the boundary $\partial\Omega$ of $\Omega$. The regularity of any $u \in W_m^l(\Omega)$ has been established in the following theorem:

**Theorem 1.1** *(cf. [21]) If $u \in W_m^l(\Omega)$ and an integer $k$ satisfies the inequality $m(l-k) > n$ then there exists a function $u \in C^k(\Omega)$ and a constant $M$ such that*

$$||u||_{C^k(\Omega)} \leq M\ ||u||_{W_m^l(\Omega)}.$$

# Chapter 2

# Variational Calculus

## 2.1 Introduction

As we know the finite difference methods for solving differential equations are applicable when a differentiable regular solution exists. Obviously, the existence of regular solutions of a differential equation depends on its coefficients. Differential equations with discontinuous coefficients may not have differentiable solutions, however, they may have weak solutions, instead. This is a restriction for the range of differential equations for which the finite difference method can be used successfully. Nevertheless, weak solutions as well as regular solutions can be approximated by variational methods or finite element methods. These methods are based on variational principles and the theory of distributions. In order to present the essence of variational and finite element methods we have given in the previous chapter a preliminary description of distributions, generalized functions and weak solutions. We shall start by introducing the idea of weak solutions using the following simple differential equation:

$$\frac{du}{dx} = \begin{cases} 1 \quad if \quad 0 \leq x \leq \dfrac{1}{2}, \\ 0 \quad if \quad \dfrac{1}{2} < x \leq 1, \end{cases} \tag{2.1}$$

with the initial value condition $u(0) = 0$.
One can check that

$$u(x) = \begin{cases} x \quad if \quad 0 \leq x \leq \dfrac{1}{2}, \\ \dfrac{1}{2} \quad if \quad \dfrac{1}{2} < x \leq 1. \end{cases}$$

satisfies the equation (2.1) at each $x \neq \frac{1}{2}$. Obviously, $u(x)$ is not differentiable at $x = \frac{1}{2}$. Thus, $u(x)$ cannot be a regular solution of equation (2.1) in the interval $[0, 1]$. However, $u(x)$ is called a weak solution of (2.1), since $u(x)$

satisfies the following variational equation:

$$\int_0^1 \frac{du(x)}{dx}\psi(x)dx = \int_0^1 f(x)\psi(x)dx,$$

or, integrating by parts

$$\int_0^1 u(x)\frac{\psi(x)}{dx}dx = -\int_0^1 f(x)\psi(x)dx,$$

for any $\psi \in C_0^\infty(0,1)$,[1] where

$$f(x) = \begin{cases} 1 \quad if \quad 0 \leq x \leq \frac{1}{2}, \\ 0 \quad if \quad \frac{1}{2} < x \leq 1, \end{cases}$$

In general, regular solutions of a differential equation are differentiable functions up to the order of a differential equation. However, weak solutions of a differential equation are distributions that satisfy relevant variational equations and they may not be differentiable at some points.

## 2.2 Variational Problems

Variational problems and equations are closely related with Euler-Lagrange-Ostrogradsky equation. We shall start with derivation of Euler-Lagrange-Ostrogradsky equation for the following functional (cf. [4]):

$$F(u) = \int_\Omega G(x, u, u_x)dx, \tag{2.2}$$

determined by a sufficiently smooth function $G(x, q, r)$ given for $x = (x_1, x_2, ..., x_n) \in \Omega$, $q \in R$ and $r = (r_1, r_2, ..., r_n) \in R^n$, where $\Omega \subset R^n$ is a bounded domain.

The variational problem to find a minimum of the functional $F(u)$ leads to Euler-Lagrange-Ostrogradsky equation. This equation is a necessary condition for the existence of a minimum of the functional $F(u)$ in a convex set. In order to find a minimum of $F(u)$, we can apply Ritz or Galerkin methods. These methods, we shall present later on. Now, let us concentrate on Euler-Lagrange-Ostrogradsky equation.

**Variational Principle 2.1** *Find a minimum of the functional $F(u)$ in the convex set $C_\phi^2(\Omega)$ of all twice continuously differentiable functions in the closed domain $\overline{\Omega}$ which satisfy the boundary condition*

$$v(x) = \phi(x), \qquad x \in \partial\Omega.$$

[1] $C_0^\infty(0.1)$ is the class of all infinitely times differentiable functions on interval $[0,1]$ which vanish together with all derivatives at $x = 0$ and $x = 1$

The equivalent statement for the problem (2.1) is:

**Variational Principle 2.2** *Find a function* $u \in C^2_\phi(\Omega)$ *which satisfies the inequality*

$$F(u) \leq F(v)$$

*for all* $v \in C^2_\phi(\Omega)$.

If $u$ is an element of the space $C^2_\phi(\Omega)$ at which $F(u)$ attains its minimum, then we shall write it as follows:

$$F(u) = \min_{v \in C^2_\phi(\Omega)} F(v).$$

**The necessary condition.** We shall obtain Euler-Lagrange-Ostrogradsky equation as a necessary condition for existence of a minimum of $F(u)$ in $C^2_\phi(\Omega)$. Let us assume that $u \in C^2_\phi(\Omega)$ is a solution of the variational problem (2.2). Then, the function

$$\Phi(t) = \int_\Omega G(x, u + t\eta, u_x + t\eta_x)dx, \qquad -\infty < t < \infty, \tag{2.3}$$

attains its minimum at $t = 0$ for any $\eta \in C^2_0(\Omega)$. Therefore

$$\frac{d\Phi(t)}{dt} = 0 \quad \text{for} \quad t = 0. \tag{2.4}$$

On the other hand

$$\frac{d\Phi(t)}{dt} = \int_\Omega [\frac{\partial G}{\partial u}\eta + \sum_{i=1}^n \frac{\partial G}{\partial r_i}\frac{\partial \eta}{\partial x_i}]dx. \tag{2.5}$$

Integrating by parts, we obtain

$$\int_\Omega \sum_{i=1}^n \frac{\partial G}{\partial r_i}\frac{\partial \eta}{\partial x_i}dx = -\int_\Omega \eta \sum_{i=1}^n \frac{\partial}{\partial x_i}(\frac{\partial G}{\partial r_i})dx. \tag{2.6}$$

Hence, by (2.4), (2.5) and (2.6), we arrive at the equation

$$\int_\Omega \eta[\frac{\partial G}{\partial u} - \sum_{i=1}^n \frac{\partial}{\partial x_i}(\frac{\partial G}{\partial r_i}]dx = 0 \tag{2.7}$$

for any $\eta \in C^2_0(\Omega)$.
Because equation (2.7) holds for each $\eta \in C^2_0(\Omega)$, and $G(x, q, r)$ is a continuously differentiable function, therefore by (2.7), we obtain the following *Euler-Lagrange-Ostrogradsky Equation*

$$\frac{\partial G}{\partial u} - \sum_{i=1}^n \frac{\partial}{\partial x_i}(\frac{\partial G}{\partial r_i})dx = 0. \tag{2.8}$$

We shall write this equation in the following explicit form:

$$\sum_{i,j=1}^n \frac{\partial^2 G}{\partial r_i \partial r_j}\frac{\partial^2 u}{\partial x_i \partial x_j} + \sum_{i=1}^n \frac{\partial^2 G}{\partial u \partial x_i}\frac{\partial u}{\partial x_i} + \sum_{i=1}^n \frac{\partial^2 G}{\partial r_i \partial x_i} - \frac{\partial G}{\partial u} = 0. \tag{2.9}$$

**Example 2.1** *Let us note that the Euler's equation of the functional*

$$F(u) = \int_a^b G(x, u, \frac{du}{dx})dx, \qquad u \in C^2_\phi(a, b).$$

*in one space variable $x$ has the following form:*

$$\frac{\partial G}{\partial u} - \frac{d}{dx}(\frac{\partial G}{\partial r}) = 0, \tag{2.10}$$

*where $G(x, q, r)$ is a given smooth function of the variables $x \in [a, b]$, $q, r \in (-\infty, \infty)$.*
*We can write this equation in the explicit form*

$$\frac{\partial G}{\partial q} - \frac{\partial^2 G}{\partial x \partial q} - \frac{\partial^2 G}{\partial q \partial r}\frac{du}{dx} - \frac{\partial^2 G}{\partial r^2}\frac{d^2 u}{dx^2} = 0.$$

**Example 2.2** *Consider the functional*

$$F(u) = \int_a^b [(\frac{\partial u}{\partial x})^2 - \sigma(x)u^2 - 2f\ u]dx$$

*Find Euler's equation of the functional $F(u)$.*

**Example 2.3** *Let us find the necessary condition for existence of a minimum of the following functional:*

$$F(u) = \int_\Omega [(\frac{\partial u}{\partial x_1})^2 + (\frac{\partial u}{\partial x_2})^2 + \sigma(x_1, x_2)u^2 - 2f(x_1, x_2)u]dx$$

*in the convex set $C^2_\phi(\Omega)$, where $\sigma(x_1, x_2) \geq 0$, $f(x_1, x_2)$ and $\phi(x_1, x_2)$ are given continuous functions in $\Omega$ and on $\partial\Omega$, respectively.*

The function

$$G(x, q, r) = r_1^2 + r_2^2 + \sigma(x)q^2 - 2f(x)q$$

is continuous with respect to the variables $x, q$ and $r = (r_1, r_2)$ and infinitely continuously differentiable with respect to the variables $q$ and $r$. Therefore, by the equation (2.9), we obtain the following Euler-Lagrange-Ostrogradsky equation:

$$-\frac{\partial^2 u}{\partial x_1^2} - \frac{\partial^2 u}{\partial x_2^2} + \sigma(x)u = f(x), \qquad x \in \Omega,$$

with the boundary value condition

$$u(x_1, x_2) = \phi(x_1, x_2), \qquad (x_1, x_2) \in \Omega.$$

Thus, any minimum of the functional $F(u)$ must be a solution of the Euler-Lagrange-Ostrogradsky equation.
The Euler-Lagrange-Ostrogradsky equation has many important applications in different areas of mathematics and physics. In this text, Euler's equation is used as a base for variational methods and finite element methods. We shall present these methods in the next sections. As an application of the Euler-Lagrange-Ostrogradsky equation in mechanics, the *Hamilton Principle* and Newton laws of motion can be derived from this equation.

## 2.3 Relationship Between Linear Differential Equations and Variational Problems

In the previous section, we have shown that if the functional $F(u)$ attains its minimum at $u$ then the function $u$ (or distribution $u$) is a solution of Euler's equation which satisfies the boundary value condition

$$u = \phi \qquad \text{on} \quad \partial\Omega.$$

As we know, Euler's equation is the necessary condition for existence of a minimum of a functional $F(u)$. For certain class of functionals, Euler's equation is also a sufficient condition for existence of the minimum of $F(u)$. This class of functionals, we present below.
Let us consider the following equation:

$$Lu = f(x), \qquad x \in \Omega, \tag{2.11}$$

where $L$ is a linear operator in the Hilbert space $H$, and a given function $f$ is an element of $H$.
We assume that the operator $L$ is positive definite in the space $H$. Thus, $L$ satisfies the conditions of the following definition:

**Definition 2.1** *A linear operator $L$ is said to be positive definite in the Hilbert space $H$ if*

1. *$L$ is a symmetric operator, and its domain $D(L)$ is dense everywhere in $H$, i.e. $(Lu, v) = (u, Lv)$, for any $u, v \in D(L)$ and $\overline{D(L)} = H$,*

2. *there exists a constant $\gamma > 0$ such that*

$$(Lu, u) \geq \gamma(u, u) \qquad \text{for all} \quad u \in D(L).$$

**Example 2.4** *Let us consider the operator $L = -\dfrac{d^2u}{dx^2}$ with its domain $D(L) = C_0^2(0, a)$. This operator is positive definite in the Hilbert space $H = L_2(0, a)$ of all functions square integrable on the interval $[0, a]$. Indeed, we know that the domain $D(L) = C_0^2(0, a)$ is dense everywhere in the space $L_2(0, a)$, i.e. $\overline{C_0^2} = L_2(0, a)$. Also, $L$ is a symmetric operator, since*

$$(Lu, v) = \int_0^a -\frac{d^2u}{dx^2} v dx = \int_0^a \frac{du}{dx}\frac{dv}{dx} dx = \int_0^a u(-\frac{d^2v}{dx^2}) dx = (u, Lv)$$

*for any $u, v \in C_0^2(0, a)$.*
*To check the last condition of definition, we note that*

$$u(x) = \int_0^x \frac{du(t)}{dt} dt$$

*Hence, by Cauchy's inequality*

$$u^2(x) = [\int_0^x \frac{du(t)}{dt} dt]^2 \leq a \int_0^a [\frac{du}{dx}]^2 dx.$$

*Integrating both sides of the above inequality, we obtain*

$$||u||^2 = \int_0^a u^2 dx \leq a^2 \int_0^a [\frac{du}{dx}]^2 dx. \tag{2.12}$$

*But*

$$\int_0^a [\frac{du}{dx}]^2 dx = -\int_0^a u\frac{du}{dx} dx = (Lu, u),$$

*so that, by (2.12)*

$$(Lu.u) \geq a^2(u, u) \qquad \text{for any} \qquad u \in C_0^2(0, a).$$

*Therefore constant* $\gamma = \dfrac{1}{a^2}$.

Let us consider the class of functionals of the following form:

$$F(u) = (Lu, u) - 2(f, u), \qquad u \in D(L), \quad f \in H, \tag{2.13}$$

where $L$ is a positive definite operator in the Hilbert space $H$.
The real number

$$\sqrt{(Lu, u)}$$

is called energy of $u$.
Now, we shall show that Euler's equation is necessary and sufficient condition for existence of a minimum in domain $D(L)$ of the functional $F(u)$, provided that $L$ is a positive definite operator.
The following theorem holds:

**Theorem 2.1** *Let $L$ be a linear operator positive definite in the Hilbert space $H$.*
*If $u \in D(L)$ is a regular solution of the linear equation*

$$Lu = f(x), \qquad x \in \Omega, \quad f \in H, \tag{2.14}$$

*then the functional $F(u)$ attains its minimum at $u \in D(L)$, i.e.,*

$$F(u) = \min_{v \in D(L)} F(v),$$

*and if the functional $F(u)$ attains its minimum at $u \in D(L)$, then $u$ is a regular solution of the linear equation*

$$Lu = f(x), \qquad x \in \Omega, \qquad f \in H.$$

**Proof.** At first, let us assume that $u \in D(L)$ is a regular solution [2] of the linear equation

$$Lu = f(x), \quad x \in \Omega, \quad f \in H.$$

Then, the following identity holds:

$$F(v) = (Lv, v) - 2(f, v) = (Lv, v) - 2(Lu, v) = (L(v-u), v-u) - (Lu, u)$$

for any $v \in D(L)$.
Hence

$$\min_{v \in D(L)} F(v) = \min_{v \in D(L)} [(L(v-u), v-u) - (Lu, u)] = -(Lu, u).$$

Therefore

$$F(u) = \min_{v \in D(L)} F(v) = -(Lu, u).$$

If $u \in D(L)$ is a solution of the linear equation (2.14), then the functional $F(u)$ attains its minimum in $D(L)$ at $u$.
Now, let us assume that the functional $F(u)$ attain its minimum at certain $u \in D(L)$, i.e.,

$$F(u) = \min_{v \in D(L)} F(v), \quad u \in D(L).$$

Then, the following inequality holds:

$$F(u) \leq F(u + t\eta)$$

for any $\eta \in D(L) \cup C_0(\Omega)$ and $t \in (-\infty.\infty)$. [3] Because the quadratic function of the real variable $t$

$$F(u + t\eta) = t^2(L\eta, \eta) + 2t[(Lu, \eta) - (f, \eta)] + F(u)$$

attains its minimum at $t = 0$, it follows that

$$\frac{dF(u + t\eta)}{dt} = (Lu - f, \eta) = 0, \quad \text{at} \quad t = 0,$$

for any $\eta \in D(L) \cup C_0(\Omega)$.
Hence $u \in D(L)$ is the regular solution of the linear equation $Lu = f, \quad f \in H$.
This ends the proof.
Let us note that $Lu = f$ is the Euler's equation for the functional $F(u)$ which is determined by a positive definite operator $L$ in the Hilbert space $H$. Also, by theorem (2.1), this equation has a regular solution $u \in D(L)$ if and only if the functional $F(u)$ attains its minimum at $u \in D(L)$.
Theorem (2.1) concerns only regular solutions of differential equations. We shall consider existence and uniqueness of weak solutions of a linear equation in the next section. Now, let us illustrate theorem (2.1) with the following example:

[2] $u$ is a regular solution of the equation $Lu = f$ if $u \in D(L)$

[3] $C_0(\Omega)$ is the set of all continuous functions in $\Omega$ which vanish at the boundary $\partial\Omega$ of $\Omega$.

**Example 2.5** *Since the operator* $L = -\dfrac{d^2}{dx^2}$ *is positive definite in the space* $H = L_2(0,a)$, *when* $D(L) = C_0^2(0,a)$, *therefore,* $u \in D(L)$ *is a regular solution of the boundary value problem*

$$-\frac{d^2u(x)}{dx^2} = f(x), \qquad f \in L_2(0,a), \qquad 0 \le x \le a,$$

$$u(0) = 0, \qquad u(a) = 0$$

*if and only if the functional*

$$F(u) = (Lu,u) - 2(f,u) = \int_0^a [(\frac{du}{dx})^2 - 2f(x)u(x)]dx$$

*attains its minimum at* $u \in D(L) = C_0^2(0,a)$.

## 2.4 Existence and Uniqueness of Weak Solutions

Let us consider the linear equation

$$Lu = f(x), \qquad x \in \Omega, \qquad f \in H, \tag{2.15}$$

where $L$ is a positive definite operator in the Hilbert space $H$.
By theorem (2.1), this equation has a regular solution $u \in D(L)$ if and only if the functional

$$F(v) = (Lv,v) - 2(f,v), \qquad v \in D(F), \quad f \in H, \tag{2.16}$$

attains its minimum at $u \in D(L)$.
At first, let us note that the domain $D(F)$ of the functional $F$ is essentially greater than the domain $D(L)$ of the operator $L$. Indeed, if $v \in D(L)$ then, evidently $v \in D(F)$, so that $D(L) \subseteq D(F)$. However, there are elements of $D(F)$ which do not belong to $D(L)$. For instance, when

$$Lv = -\frac{d^2v}{dx^2} \quad \text{and} \quad D(L) = C_0^2(0,a), \tag{2.17}$$

the functional

$$F(v) = (Lv,v) - 2(f,v) = \int_0^a [-\frac{d^2v}{dx^2}v - 2fv]dx = \int_0^a [(\frac{dv}{dx})^2 - 2fv]dx \tag{2.18}$$

is well defined at all $v \in D(F) = W_2^{01}(0,a)$ [4]. However, there are functions in the Sobolev space $W_2^{01}(0,a)$ which are not twice differentiable in the interval $[0,a]$. Therefore, $D(L)$ is a proper subset of $D(F)$, i.e., $D(L) \subset D(F)$.

[4] $W_2^{01}(0,a)$ is the space of all functions which have first derivatives integrable with square and vanish at the ends of the interval $[0,a]$

Nevertheless, the domain $D(L)$ can be enlarged up to the domain $D(F)$ by the so-called Fridrich's closure of the set $D(L)$. By the assumption, the operator $L$ is positive definite in the Hilbert space $H$. We can therefore introduce the new inner product

$$(u,v)_L = (Lu,v), \qquad \text{for any} \quad u,v \in D(L),$$

in the domain $D(L) \subset H$, determined by the operator $L$. Then, the new norm

$$||u||_L = \sqrt{(Lu,u)}, \qquad u \in D(L).$$

Now, we shall consider the closure $\overline{D(L)}$ of the domain $D(L)$ in the norm $||-||_L$. This closure is the new Hilbert space

$$H_L = \overline{D(L)} \subseteq H$$

in which the functional $F(u)$ is well definite. The domain $D(F)$ of the functional $F(u)$ equals $H_L$, i.e., $D(F) = H_L$. In the literature, $H_L$ is called as the Fridrich's space or energetic space, since the norm $||u||_L$ is the energy of $u$.

**Example 2.6** *As we know, the differential operator*

$$Lv = -\frac{d^2v}{dx^2}, \qquad v \in D(L) = C_0^2(0,a)$$

*is positive definite in the space $L_2(0,a)$. Therefore, the inner product*

$$(u,v)_L = (Lu,v), \qquad u,v \in C_0^2(0,a)$$

*determines energy of $u$ as the norm*

$$||u||_L = \sqrt{(Lu,u)} = \sqrt{\int_0^a (\frac{du}{dx})^2 dx}, \qquad u \in C_0^2(0,a).$$

*Thus, the closure of the domain $C_0^2(0,a)$ in the norm $||-||_L$ implies the Sobolev space $W_2^{01}(0,a)$, i.e., then the energetic space*

$$H_L = \overline{D(L)} = \overline{C_0^2(0,a)} = W_2^{01}(0,a).$$

*Let us observe that $||-||_L$ is the norm in the energetic space $H_L = W_2^{01}(0,a)$, and it is a pseudo-norm in the Hilbert's space $H = L_2(0,a)$,* [5]

Below, we shall state and prove the fundamental theorem of variational calculus on existence and uniqueness of extremals of a functional.

[5] note that, the norm $||c||_L$ of a constant $c$ is equal to zero, but, if $c \in W_2^{01}$, then $c = 0$.

**Theorem 2.2** *If operator $L$ is positive definite in the Hilbert space $H$, then there exists a unique element $u_0 \in H_L$ at which the functional*

$$F(u) = (Lu, u) - 2(f, u), \qquad f \in H,$$

*attains its minimum in the energetic space $H_L$, and*

$$F(u_0) = \min_{v \in H_L} F(v) = -(u_0, u_0)_L = -(Lu_0, u_0).$$

**Proof.** By the assumption, the operator $L$ is positive definite in the Hilbert space $H$. Therefore, there exists a constant $\gamma > 0$ such that

$$(u, u)_L = (Lu, u) \geq \gamma(u, u), \quad \text{for any} \quad u \in D(L).$$

Hence, by Cauchy's inequality

$$|(f, u)| \leq ||f|| \, ||u|| \leq \frac{||f||}{\sqrt{\gamma}} ||u||_L \quad \text{for any} \quad u \in H_L.$$

The linear functional $(f, u)$ is continuous in the Hilbert space $H_L$. Therefore, by Riesz theorem, there exists an element $u_0 \in H_L$ such that

$$(f, u) = (u_0, u)_L \quad \text{for any} \quad u \in H_L, \quad f \in H.$$

Then, we have

$$F(u) = (u, u)_L - 2(u_0, u)_L = (u - u_0, u - u_0)_L - (u_0, u_0)_L, \qquad u \in H_L. \tag{2.19}$$

Hence, the functional $F(u)$ attains its minimum at $u_0 \in H_L$, i.e.,

$$\min_{v \in H_L} F(v) = -(u_0, u_0)_L = F(u_0).$$

Now, let us show that $u_0$ is a unique element in $H_L$ at which the functional $F(u)$ attains its minimum. Let $u_1 \in H_L$ be also an element at which $F(u)$ attains its minimum in $H_L$. Then, we have

$$F(u_0) \leq F(u_1) \quad \text{and} \quad F(u_1) \leq F(u_0)$$

This means that

$$F(u_0) = F(u_1).$$

On the other hand, by (2.19)

$$F(u_1) = (u_1 - u_0, u_1 - u_0) - (u_0, u_0)_L = (u_1 - u_0, u_1 - u_0) - F(u_0).$$

Hence $(u_1 - u_0, u_1 - u_0) = 0$ and $u_1 = u_0$. This ends the proof.
*The element $u_0 \in H_L$ at which the functional $F(u)$ attains its minimum in the energetic space $H_L$ is called a weak solution of the linear equation*

$$Lu = f(x), \qquad x \in \Omega, \quad f \in H.$$

*If $u_0 \in D(L)$ then $u_0$ becomes a regular solution of this equation.*

**Example 2.7** *Let us consider the following equation:*

$$Lu \equiv -\frac{d^2u(x)}{dx^2} = \begin{cases} 0 & if \quad 0 \le x \le \frac{1}{2}, \\ -4 & if \quad \frac{1}{2} < x \le 1, \end{cases} \tag{2.20}$$

*with the homogeneous boundary conditions* $u(0) = 0, \quad u(1) = 0.$

As we know, the operator $L$ is positive definite in the space $H = L_2(0,1)$ and the energetic space $H_L = W_2^{01}(0,1)$. Therefore, by theorem (2.2) the functional

$$F(u) = (Lu, u) - 2(f, u) = \int_0^1 [(\frac{du(x)}{dx})^2 - 2f(x)u(x)]dx, \qquad u \in H_L,$$

$$f(x) = \begin{cases} 0 & if \quad 0 \le x \le \frac{1}{2}, \\ -4 & if \quad \frac{1}{2} < x \le 1, \end{cases}$$

attains its minimum in $W_2^{01}(0,1)$ at the unique function

$$u_0(x) = \begin{cases} -\frac{1}{2}x & if \quad 0 \le x \le \frac{1}{2}, \\ 2x^2 - \frac{5}{2}x + \frac{1}{2} & if \quad \frac{1}{2} < x \le 1. \end{cases} \tag{2.21}$$

Then, the value of this minimum is:

$$F(u_0) = -(u_0.u_0)_L = -\int_0^1 f(x)u_0(x)dx = -\frac{1}{12}.$$

So that, $u_0 \in W_2^{01}(0,1)$ is the unique weak solution of the differential equation (2.20), but $u_0$ is not a regular solution of this equation, since $u_0(x)$ is not twice differentiable at $x = \frac{1}{2}$.

## 2.5 Variational Equations

In this section, we shall consider variational equations associated with positive definite operators. We shall then show that such equations are equivalent with relevant variational problems. Let $L$ be a positive definite operator in the Hilbert space $H$. Multiplying the linear equation

$$Lu = f(x), \qquad x \in \Omega, \qquad f \in H, \tag{2.22}$$

by $\eta \in H_L$, we obtain the variational form of this equation

$$(Lu, \eta) = (f, \eta), \qquad f \in H, \tag{2.23}$$

for any $\eta \in H_L$.
The following theorem holds:

**Theorem 2.3** *If $L$ is a positive definite operator in the Hilbert space $H$, then there exists a unique solution $u_0 \in H_L$ of the variational equation*

$$(Lu, \eta) = (f, \eta), \qquad \text{for any} \quad \eta \in H_L \tag{2.24}$$

*at which the functional*

$$F(u) = (Lu, u) - 2(f, u), \qquad f \in H,$$

*attains its minimum in the energetic space $H_L$.*

**Proof.** By theorem (2.2), there exists a unique element $u_0 \in H_L$ at which the functional $F(u)$ attains its minimum in the energetic space $H_L$, so that

$$F(u_0 + t\eta) \leq F(u_0)$$

for any $\eta \in H_L, \quad -\infty < t < \infty$.
Since

$$\begin{aligned} F(u_0 + t\eta) &= (L(u_0 + t\eta), u_0 + t\eta) - 2(f u_0 + t\eta) \\ &= (L\eta, \eta)t^2 + 2[(Lu_0, \eta) - (f, \eta)]t + F(u_0) \geq F(u_0), \end{aligned}$$

it follows that

$$(L\eta, \eta)t^2 + 2[(Lu_0, \eta) - (f, \eta)]t \geq 0$$

for any $\eta \in H_L$ and $t \in (-\infty, \infty)$.
Hence

$$(Lu_0, \eta) = (f, \eta) \qquad \text{for any} \quad \eta \in H_L.$$

To complete the proof, we should show that at each solution $u_0 \in H_L$ of (2.24), the functional $F(u)$ attains its minimum in the space $H_L$. Since $L$ is a positive definite operator in the Hilbert space $H$, we have

$$F(\eta) - F(u_0) = F(u_0 + (\eta - u_0)) - F(u_0) = (L(\eta - u_0), \eta - u_0) \geq 0$$

for any $\eta \in H_L$.
Hence

$$F(u_0) \leq F(\eta) \qquad \text{for any} \quad \eta \in H_L,$$

and

$$F(u_0) = \min_{v \in H_L} F(v).$$

**Example 2.8** *Let us recall equation (2.20)*

$$Lu \equiv -\frac{d^2u(x)}{dx^2} = \begin{cases} 0 & if \quad 0 \leq x \leq \frac{1}{2}, \\ -4 & if \quad \frac{1}{2} < x \leq 1, \end{cases} \tag{2.25}$$

*with the homogeneous boundary conditions* $u(0) = 0, \quad u(1) = 0.$
*The variational form of this equation is:*

$$-\int_0^1 \frac{d^2u(x)}{dx^2}\eta(x)\,dx = \int_0^1 \frac{du(x)}{dx}\frac{\eta(x)}{dx}dx = -4\int_{\frac{1}{2}}^1 \eta(x)dx$$

*for any* $\eta \in H_L = W_2^{01}(0.1)$.
*By theorem (2.3), this equation has the unique weak solution* $u_0 \in H_L$ *given by the formula (2.21). Indeed*

$$-\int_0^1 \frac{d^2u(x)}{dx^2}\eta(x)\,dx = \int_0^1 \frac{du_0}{dx}\frac{d\eta}{dx}dx = -\frac{1}{2}\int_0^{\frac{1}{2}} \frac{d\eta}{dx}dx + \int_{\frac{1}{2}}^1 (4x - \frac{5}{2})\frac{d\eta}{dx}dx = -4\int_{\frac{1}{2}}^1 \eta dx.$$

*for any* $\eta \in W_2^{01}(0,1)$.

## 2.6 Ritz and Galerkin Methods

**The Ritz's Method.** We shall apply Ritz's method to find a minimum of the functional

$$F(v) = (Lv, v) - 2(f, v), \qquad v \in H_L, \quad f \in H, \tag{2.26}$$

in the energetic space $H_L$, where $L$ is a positive definite operator in the Hilbert's space $H$.
First, we choose a complete set of elements in the space $H_L$,

$$\phi_1, \phi_2, ..., \phi_n, ...; \qquad \phi_i \in H_L, \ \ i = 1.2....;$$

The elements $\phi_i$, $1, 2, ...;$ are called the Ritz's coordinates in the space $H_L$. We can approximate an element $u \in H_L$ by a linear combination

$$a_1\phi_1 + a_2\phi_2 + \cdots + a_N\phi_N$$

with a given accuracy $\epsilon > 0$, so that for every $\epsilon > 0$ there exists an integer $N$ such that

$$||u - (a_1\phi_1 + a_2\phi_2 + \cdots + a_N\phi_N)||_L < \epsilon.$$

Next, we find

$$u_N = a_1\phi_1 + a_2\phi_2 + \cdots + a_N\phi_N$$

at which the functional $F(v)$ attains minimum in the subspace

$$X_N = span\{\phi_1, \phi_2, ..., \phi_N\}$$

of the space $H_L$. Obviously, $u_N \in X_N$ approximates the element $u \in H_L$ with the error $||u_N - u||_L$.

In order to determine the coefficients $a_1, a_2, ..., a_N$, we consider the following function:

$$\begin{aligned}\Phi(a_1, a_2, ...a_N) &= F(a_1\phi_1 + a_2\phi_2 + \cdots + a_N\phi_N) \\ &= \sum_{i=1}^{N}\sum_{k=1}^{N}(L\phi_i, \phi_k)a_i a_k - 2\sum_{i=1}^{N}(f, \phi_i)a_i.\end{aligned}$$

This quadratic function has its minimum at a point $a_1^*, a_2^*, ..., a_N^*$ if and only if

$$\frac{\partial\Phi(a_1^*, a_2^*, ..., a_N^*)}{\partial a_k} = 0, \quad \text{for} \quad k = 1, 2, ..., N. \tag{2.27}$$

On the other hand

$$\frac{\partial\Phi(a_1, a_2, ...a_N)}{\partial a_k} = 2\sum_{i=1}^{N}(L\phi_i, \phi_k)a_i - 2(f, \phi_k), \quad \text{for} \quad k = 1, 2, ..., N. \tag{2.28}$$

Comparing (2.27) and (2.28), we arrive at the following Ritz's system of linear equations:

$$\begin{aligned}
&(L\phi_1, \phi_1)a_1 + (L\phi_2, \phi_1)a_2 + \cdots + (L\phi_N, \phi_1)_L a_N = (f, \phi_1) \\
&(L\phi_1, \phi_2)a_1 + (L\phi_2, \phi_2)a_2 + \cdots + (L\phi_N, \phi_2)a_N = (f, \phi_2) \\
&(L\phi_1, \phi_3)a_1 + (L\phi_3, \phi_1)a_2 + \cdots + (L\phi_N, \phi_3)a_N = (f, \phi_3) \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
&(L\phi_1, \phi_N)a_1 + (L\phi_2, \phi_N)a_2 + \cdots + (L\phi_N, \phi_N)a_N = (f, \phi_N)
\end{aligned} \tag{2.29}$$

Let us note that the matrix $A = \{(\phi_i, \phi_k)\},\ i, k = 1, 2, ..., N$ of Ritz's system of equations (2.29) is non-singular, since it is a Gram's matrix of linearly independent elements $\phi_1, \phi_2, ..., \phi_N$. Therefore, this system of equations has the unique solution $a_1^*, a_2^*, ..., a_N^*$ which uniquely determines $u_N \in X_N \subset H_L$.

**Example 2.9** *let us recall equation (2.7)*

$$Lu \equiv -\frac{d^2u(x)}{dx^2} = \begin{cases} 0 & if \quad 0 \le x \le \frac{1}{2}, \\ -4 & if \quad \frac{1}{2} < x \le 1, \end{cases} \tag{2.30}$$

*with the homogeneous boundary conditions* $u(0) = 0, \quad u(1) = 0.$

As we know, in this example the energetic space $H_L = W_2^{01}(0,1)$. We can choose the set of linearly independent functions in the space $W_2^{01}(0,.1)$ as follows:

$$\phi_k(x) = \sin\ k\pi x, \qquad 0 \le x \le 1, \quad k = 1,2,...$$

So that

$$(L\phi_k, \phi_l) = k^2\pi^2 \int_0^1 \sin k\pi x \sin l\pi x dx = \begin{cases} \dfrac{k^2\pi^2}{2} & if \quad k = l, \\ 0 & if \quad k \neq l, \end{cases} \tag{2.31}$$

and

$$(f, \phi_k) = -4\int_{\frac{1}{2}}^1 \sin k\ \pi x\ dx = \frac{4}{k\pi} \begin{cases} -1 & if \quad k = 1,3,5,7,...; \\ 2 & if \quad k = 2,6,10,14,...; \\ 0 & if \quad k = 4,8,12,16,...; \end{cases} \tag{2.32}$$

Whence

$$a_k^* = \frac{8}{k^3\pi^3} \begin{cases} -1 & if \quad k = 1,3,5,7,...; \\ 2 & if \quad k = 2,6,10,14,...; \\ 0 & if \quad k = 4,8,12,16,...; \end{cases} \tag{2.33}$$

Thus, the Ritz's approximation of the weak solution $u_0(x)$ is:

$$\begin{aligned} u_N(x) = \quad & \frac{8}{\pi^3}[-\sin \pi x + \frac{2}{2^3}\sin 2\pi x - \frac{1}{3^3}\sin 3\pi x - \\ & -\frac{1}{5^3}\sin 5\pi x + \frac{2}{6^3}\sin 6\pi x - \frac{1}{7^3}\sin 7\pi x - \frac{1}{9^3}\sin 9\pi x + \\ & +\frac{2}{10^3}\sin 10\pi x - \frac{1}{11^3}\sin 11\pi x - \cdots + a_N^*\sin N\pi x]. \end{aligned} \tag{2.34}$$

Let us note that the function $u_N(x)$ given by formula (2.34) is the partial sum of the Fourier series of the weak solution $u_0(x)$.
In order to solve the equation

$$-\frac{d^2u(x)}{dx^2} = f(x), \qquad a \le x \le b \tag{2.35}$$

with the homogeneous boundary value conditions for a continuous function $f(x)$, we can use the following *Mathematica* module

**Program 2.1** *Mathematica module that solves the equation (2.35).*

---

```
    ritz[f_,psi_,n_,a_,b_]:=
         Module[{ },
           m=Table[Integrate[D[psi[x,i],x]*D[psi[x,k],x],{x,a,b}],
           {i,1,n},{k,1,n}];
           p=Table[Integrate[f[x]*psi[x,i],{x,a,b}],{i,1,n}];
           c=LinearSolve[m,p];
           Print["u(x,",n,") = ",Sum[c[[i]]*psi[x,i],{i,1,n}]]
 ]
```

---

By executing the following instructions

```
f[x_]:=-4;
psi[x_,i_]:=Sin[Pi*x*i];
ritz[f,psi,8,0,1];
```

we obtain the solution [6]

$$
\begin{aligned}
u(x,8) = \;& -\frac{8\sin\pi x}{\pi^3}+\frac{2\sin 2\pi x}{\pi^3}-\frac{8\sin 3\pi x}{27\pi^3} \\
& -\frac{8\sin 5\pi x}{125\pi^3}+\frac{2\sin 6\pi x}{27\pi^3}-\frac{8\sin 7\pi x}{343\pi^3}-\frac{8\sin 9\pi x}{729\pi^3}.
\end{aligned}
\tag{2.36}
$$

**The Galerkin's Method.** We can apply the Galerkin's method to solve the following variational equation:

$$(Lu,\eta)=(f,\eta), \qquad f\in H, \tag{2.37}$$

for any $\eta\in H_L$, where $L$ is a positive definite operator in the Hilbert space $H$. Similarly as for the Ritz's method, at first, we choose in the energetic space $H_L$ a complete set[7] of elements

$$\phi_1,\phi_2,...,\phi_n,...; \qquad \phi_i\in H_L, \;\; i=1.2....;$$

Thus, we can approximate any element $u\in H_L$ by a linear combination

$$a_1\phi_1+a_2\phi_2+\cdots+a_N\phi_N$$

with a given accuracy $\epsilon>0$, i.e., for every $\epsilon>0$ there exists an integer $N$ such that

$$||u-(a_1\phi_1+a_2\phi_2+\cdots+a_N\phi_N)||_L<\epsilon.$$

Then, the variational equation (2.37) is equivalent to the following:

$$(Lu,\phi_k)=(f,\phi_k), \qquad k=1,2,...; \tag{2.38}$$

[6] Note that in module, we set 1/2 instead of a, in the evaluation of right sides p, since $f(x)=0$ for $0\le x\le 1/2$.

[7] It is sufficient to choose a set which determines such subspace $X\subseteq H_L$ that the solution $u\in X$

In order to approximate the solution $u \in X \subseteq H_L$, we consider the subspace

$$X_N = span\{\phi_1, \phi_2, ..., \phi_N\}$$

of the space $X$. Then, we find

$$u_N = a_1\phi_1 + a_2\phi_2 + \cdots + a_N\phi_N \in X_N$$

which satisfies the equations

$$(Lu_N, \phi_k) = (f, \phi_k), \qquad k = 1, 2, ..., N,$$

or

$$\sum_{i=1}^{N}(L\phi_i, \phi_k)a_i = (f, \phi_k), \qquad k = 1, 2, ..., N \tag{2.39}$$

Let us observe that this system of linear equations is the same as the Ritz's system (2.29). Both Ritz's and Galerkin's method produce the same approximate solution $u_N$, provided that the operator $L$ is positive definite in the Hilbert space $H$.

Now, let us estimate the error of the Ritz's and Galerkin's methods.

**Error Bound.** Again, we assume that the operator $L$ is positive definite in the Hilbert's space $H$. Then, both Ritz's and Galerkin's methods determine the same approximate solution $u_N$ of the equation

$$Lu = f, \qquad f \in H, \tag{2.40}$$

so that

$$(Lu_N, \eta) = (f, \eta) \tag{2.41}$$

for any $\eta \in X_N \subset X \subseteq H_L$,

The following theorem holds:

**Theorem 2.4** *If $u$ is a solution of the equation (2.40) and $u_N$ is a solution of the equation (2.41), then the error $u_N - u$ satisfies the inequality*

$$||u_N - u||_L \leq ||\eta - u||_L \tag{2.42}$$

*for any $\eta \in X_N$, and*

$$\min_{\eta \in X_N} ||\eta - u||_L = ||u_N - u||_L. \tag{2.43}$$

**Proof.** Since

$$F(\eta) = (\eta - u, \eta - u)_L - (u, u)_L$$

for any $\eta \in H_L$, therefore

$$\min_{\eta \in X_N} F(\eta) = \min_{\eta \in X_N} (\eta - u, \eta - u)_L - (u, u)_L.$$

On the other hand

$$\min_{\eta \in X_N} F(\eta) = F(u_N) = (u_N - u, u_N - u)_L - (u, u)_L,$$

and

$$\min_{\eta \in X_N} ||\eta - u||_L = ||u_N - u||_L$$

and

$$||u_N - u||_L \leq ||\eta - u||_L,$$

for any $\eta \in X_N$.

## 2.7 Exercises

**Question 2.1** *Let $\chi(x)$ be characteristic function of the interval $[a, b]$, i.e.,*

$$\chi(x) = \begin{cases} 1 & if \quad x \in [a, b], \\ 0 & if \quad x \notin [a, b]. \end{cases}$$

*Define a distribution $F_\chi(u), \quad u \in C_0^\infty(R)$ and check that this distribution satisfies definition (1.1). Find all derivatives of the distribution $F_\chi(u)$.*

**Question 2.2** *Let $E(x)$ be the step-function, i.e., $E(x)$ is the greatest integer not greater than $x$. Find all derivatives of the distribution $F_E(u)$.*

**Question 2.3** *State the following initial value problem in terms of distributions:*

$$\frac{du(x)}{dx} = \begin{cases} 1 & if \quad 0 \leq x \leq 1, \\ 0 & if \quad 1 < x \leq 2, \end{cases} \qquad u(0) = 0. \tag{2.44}$$

*Find a weak solution of the initial value problem (2.44).*

**Question 2.4** *Solve Euler's equation of the functionals*

1. (a) $F(u) = \int_1^2 [(\frac{du}{dx})^2 + 2u\frac{du}{dx} + u^2]dx$ *when* $u(1) = 1, \quad u(2) = 0,$

   (b) $F(u) = \int_0^\pi [4u \cos x + (\frac{du}{dx})^2 - u^2]dx$ *when* $u(0) = 0, \quad u(\pi) = 0.$

*Determine the minimum of functionals (a) and (b) under given boundary value conditions.*

**Question 2.5** *Find Euler's equation for the following functionals:*

1. (a) $F(u) = \int_\Omega [(\frac{\partial u}{\partial x_1})^2 + (\frac{\partial u}{\partial x_2})^2] dx_1 dx_2, \qquad u \in W_2^{01}(\Omega).$

$$(b)\ F(u) = [2(\frac{\partial u}{\partial x_1})^2 + 3(\frac{\partial u}{\partial x_2})^2]dx_1dx_2, \qquad u \in W_2^{01}(\Omega).$$

*where* $\Omega = \{x = (x_1, x_2)\ :\ \ 0 < x_1, x_2 < 1\}$.

**Question 2.6** *Assume that, in polar coordinates, $u(\rho, \theta)$ minimizes the functional*

$$F(u) = \int_{\Omega'} [(\frac{\partial u}{\partial \rho})^2 + \frac{1}{\rho^2}\frac{\partial u}{\theta})^2]\rho\ d\rho d\theta.$$

*Show that $u(\rho, \theta)$ satisfies the equation*

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho}\frac{\partial u}{\partial \rho} + \frac{1}{\rho^2}\frac{\partial^2 u}{\partial \theta^2} = 0.$$

**Question 2.7** *Show that the following operators are positive definite in the indicated spaces:*

1. (a) $Lu \equiv -\frac{\partial^2 u}{\partial x_1^2} - \frac{\partial^2 u}{\partial x_2^2}, \qquad u \in W_2^{01}(\Omega),$

   (b) $Lu \equiv \frac{\partial^4 u}{\partial x_1^4} + 2\frac{\partial^4 u}{\partial^2 x_1^2 \partial x_2^2} + \frac{\partial^4 u}{\partial x_2^4}, \qquad u \in W_2^{02}(\Omega).$

*where* $\Omega = \{x = (x_1, x_2)\ :\ \ 0 < x_1, x_2 < 1\}$.

**Question 2.8** *Show that the boundary value problem*

$$-\frac{d^2 u}{dx^2} = sign\ x, \qquad -1 < x < 1,$$
$$u(-1) = 0, \quad u(1) = 0,$$

*has a unique weak solution in the Sobolev space $W_2^{01}(-1, 1)$.*
*Find the weak solution u of this boundary problem.*

**Question 2.9** *Solve the boundary value problem*

$$-\frac{d^2 u}{dx^2} = \begin{cases} 1 & if \quad 0 \le x \le 1, \\ 2 & if \quad 1 < x \le 2, \end{cases}$$

$$u(0) = 0 \quad \text{ and } \quad u(2) = 0,$$

*by Ritz's method using Ritz's coordinates*

$$\phi_k(x) = \sin\frac{k\pi x}{2}, \qquad k = 1, 2, ...; \qquad 0 \le x \le 2.$$

**Question 2.10** *Consider the following boundary value problem:*

$$-\frac{\partial^2 u}{\partial x_1^2} - \frac{\partial^2 u}{\partial x_2^2} = 1, \qquad (x_1, x_2) \in \Omega,$$
$$u(x_1, x_2) = 0, \qquad (x_1, x_2) \in \partial\Omega,$$

*where* $\Omega = \{x = (x_1, x_2) \ : \ 0 < x_1, x_2 < 1\}$.

1. *Use Ritz's method to find* $u_2 \in X_2$ *which approximates the solution* $u \in W_2^{01}(\Omega)$, *where the subspace*

$$X_2 = span\{\phi_1, \phi_2\} \subset W_2^{01}(\Omega),$$

$$\phi_1(x_1, x_2) = x_1 x_2 (1 - x_1)(1 - x_2), \quad \phi_2(x_1, x_2) = x_1 x_2 (1 - x_1^2)(1 - x_2^2).$$

2. *Solve this boundary value problem by the Ritz's method using the following Ritz's coordinates:*

$$\phi_{km}(x_1, x_2) = \frac{2}{\pi\sqrt{k^2 + m^2}} \sin\ k\pi x_1 \sin m\pi x_2, \qquad k, m = 1.2, ...;$$

# Chapter 3

# Polynomial Splines

## 3.1 Space $S_m(\Delta, k)$.

Polynomial splines of the class $S_m(\Delta, k)$ are successfully applied in the theory of approximation of functions as well as in solving of problems which arise in the fields of differential equations and engineering.
In order to introduce the definition of polynomial splines of degree $m$, let us first define normal partition $\Delta$ of the interval $[a, b]$.
A partition

$$\Delta \ : \quad a = x_0 < x_1 < \cdots < x_N = b,$$

is normal if there exits constant $\sigma$ such that

$$\frac{\max_{0 \le i \le N-1}(x_{i+1} - x_i)}{\min_{0 \le i \le N-1}(x_{i+1} - x_i)} = \sigma_N,$$

and $\sigma_N \le \sigma$ for all natural $N$.

**Definition 3.1** *A function $s(x)$ is said to be a polynomial spline of degree $m$ if the following conditions are satisfied:*

- *$s(x)$ is a polynomial of degree at most $m$ on each subinterval $[x_i, x_{i+1}]$, $i = 0, 1, \ldots, N-1$.*
- *$s(x)$ and its first $m-1$ derivatives are continuous functions on the interval $[a, b]$.*

*The class of all polynomial splines of degree $m$ spanned over the partition $\Delta$ shall be denoted by the symbol $S_m(\Delta, m-1)$.*

**The Basis.** Now, we shall determine a basis of the space $S_m(\Delta, m-1)$. Let us consider the following auxiliary function:

$$(x-t)_+^m = \begin{cases} (x-t)^m & if \quad x \le t, \\ 0 & if \quad x > t. \end{cases}$$

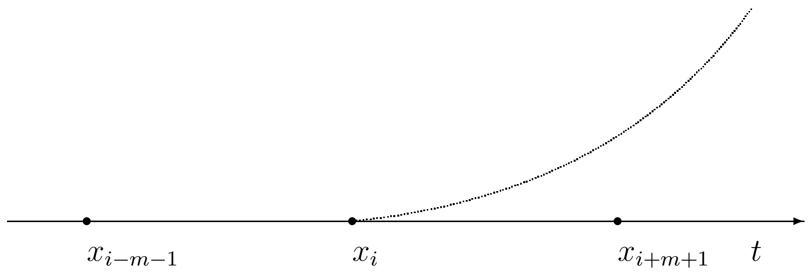



Fig. 4.1. $(x_i - t)_+^m$

The finite difference of order $m+1$ of the auxiliary function is

$$\Delta^{m+1}(x_i - t)_+^m = (-1)^{m-1}\sum_{\nu=0}^{m+1}(-1)^\nu \binom{m+1}{\nu}(x_{i+\nu} - t)_+^m.$$

**Example 3.1** *We compute the differences*

$$\begin{aligned}
m=1,\quad \Delta^2(x_i - t)_+ &= (x_i - t)_+ - 2(x_{i+1} - t)_+ + (x_{i+2} - t)_+ \\
m=2,\quad \Delta^3(x_i - t)_+^2 &= -(x_i - t)_+^2 + 3(x_{i+1} - t)_+^2 - 3(x_{i+2} - t)_+^2 \\
&\quad -(x_{i+3} - t)_+^2 \\
m=3,\quad \Delta^4(x_i - t)_+^3 &= (x_i - t)_+^3 - 3(x_{i+1} - t)_+^3 + 6(x_{i+2} - t)_+^3 \\
&\quad -3(x_{i+3} - t)_+^3 + (x_{i+4} - t)_+^3
\end{aligned}$$

Assuming that $\Delta$ is a uniform partition of interval $[a, b]$, so that

$$x_i = a + ih, \qquad i = 0, 1, ..., N; \qquad h = \frac{b-a}{N}.$$

we find the difference

$$\Delta^{m+1}(x_i - t)_+^m = 0 \quad \text{for} \quad t \geq x_i.$$

Indeed, the function

$$(x_{i+\nu} - t)_+^m = (x_{i+\nu} - t)^m, \quad \text{for} \ \ t \geq x_{i+\nu}, \quad \nu = 0, 1, ..., m+1$$

is the polynomial of degree $m$, therefore its difference of order $m+1$ is equal to zero.
Also, we note that

$$\Delta^{m+1}(x_i - t)_+^m = 0 \quad \text{for} \quad t \leq x_i,$$

Because

$$(x_{i+\nu} - t)_+^{m+1} = 0, \qquad t \leq x_i$$

for $\nu = 0, 1, .., m+1$.

Hence, we have

$$\Delta^{m+1}(x_i - t)^m_+ = \tag{3.1}$$

$$\begin{cases} \sum_{\nu=0}^{m+1} (-1)^{\nu+m+1} \binom{m+1}{\nu} (x_{i+\nu} - t)^m_+, & x_{i+q-1} < t < x_{i+q}, \quad q = 1, 2, ..., m+1, \\ 0, & t \le x_i \qquad \text{or} \quad t > x_{i+m+1} \end{cases}$$

It is clear, by (3.1), that the functions

$$K_m(x_i, t) = \Delta^{m+1}(x_i - t)^m_+, \qquad i = 0, 1, ..., N, \tag{3.2}$$

are polynomials of degree $m$ on each subinterval $[x_i, x_{i+1}]$, and they are $m-1$ times continuously differentiable in the interval $[a, b]$, i.e., $K_m(x_i, t) \in C^{m-1}[a, b]$. Therefore, each $K_m(x_i, t)$ is a polynomial spline of the class $S_m(\Delta, m-1)$. Obviously, $K_m(x_i, t)$ can be considered on the whole real line with infinite number of knots $x_i$, $i = 0, \pm 1, \pm 2, ...$; However, only $N + m$ of them are not identically equal to zero on the interval $[a, b]$. These non-zero and linearly independent splines are:

$$K_m(x_{-m}, t),\ K_m(x_{-m+1}, t),\ K_m(x_{-m+2}, t),\ \cdots, K_m(x_{N-1}, t).$$

**Example 3.2** *Let $m = 1$, then, by (3.2), we obtain*

$$K_1(i, t) = \begin{cases} 0, & t \le i, \quad \text{or} \quad t \ge i+2, \\ -i + t, & i \le t \le i+1, \\ i - t + 2, & i+1 \le t \le i+2. \end{cases}$$

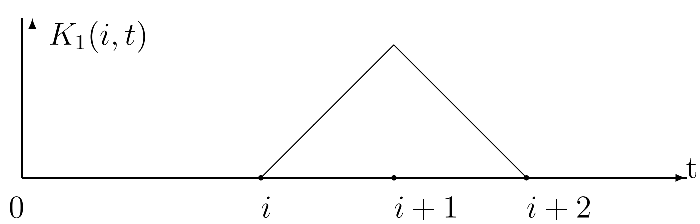



Fig 4.2. Linear spline $K_1(i, t)$

For the uniform partion of the interval $[a, b]$, we shall consider normalized splines

$$\frac{1}{h^m} K_m(x_{i-m+1}, x), \qquad i = 1, 2, ..., N+3, \quad a \le x \le b,$$

as a basis of the space $S_m(\Delta, m-1)$ of the dimension

$$dim(S_m(\Delta, m-1)) = N + m.$$

## 3.2 Properties of Splines.

**Minimum property.** Below, we shall consider the space $S_m(\Delta, m-1)$ when $m = 2q+1$ is an odd positive integer. Splines from this space minimize the following functional

$$F(g) = \int_a^b [g^{(q+1)}(x)]^2 dx, \qquad g \in C^{(q+1)}[a,b].$$

Precisely, let us consider the following variational problem (cf. [15], [20]):
**Variational Problem.** *Find a function $s \in C^{(q+1)}[a,b]$ at which the functional $F(g)$ attains its minimum under the interpolation conditions:*

$$g(x_i) = f(x_i), \qquad i = 0, 1, ..., N,$$

and either

$$g^{(j)}(a) = g^{(j)}(b) = 0, \qquad j = q+1, q+2, ..., 2q, \tag{3.3}$$

or

$$g^{(j)}(a) = g^{(j)}(b) = 0, \qquad j = 1, 2, ..., q,$$

*for a given function $f(x)$, $x \in [a,b]$.*
[1] The following theorem holds:

**Theorem 3.1** *There exists a unique spline $s \in S_m(\Delta, m-1)$ which satisfies the interpolation conditions (3.3).*

**Proof.** Every spline $s \in S_m(\Delta, m-1)$ can be written as the following linear combination:

$$s(x) = a_{-m}K_m(x_{-m}, x) + a_{-m+1}K_{-m+1}(x_{m+1}, x) + \cdots + a_{N-1}K_m(x_{N-1}, x).$$

By the interpolation conditions, $N+m$ coefficients $a_{-m}, a_{-m+1}, ..., a_{N-1}$ must satisfy the following system of $N+m$ linear equations:

$$s(x_i) = f(x_i), \quad i = 0, 1, ..., N$$

$$s^{(j)}(a) = s^{(j)}(b) = 0, \qquad j = q+1, q+2, ..., 2q, \quad m = 2q+1.$$

Clearly, there exists a unique interpolating spline $s \in S_m(\Delta, m-1)$ if the corresponding system of homogeneous equations possesses only trivial solution. To show this, we assume that $f(x_i) = 0$, for $i = 0, 1, ..., N$.

[1] The conditions put on the derivatives can be replaced by periodicity conditions.

Then, we have

$$\begin{aligned} F(s) &= \int_a^b [s^{(q+1)}(x)]^2 dx \\ &= -\int_a^b s^{(q)}(x)s^{(q+2)}(x)dx \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &= (-1)^q \int_a^b s'(x)s^{(2q+1)}(x)dx \\ &(-1)^q \sum_{i=0}^{N-1} \{s(x_{i+1})s^{(2k+1)}(x_{i+1}) - s(x_i)s^{(2k+1)}(x_i) \\ &- \int_{x_i}^{x_{i+1}} s(x)s^{(2q+2)}(x)dx\}. \end{aligned}$$

Hence, $F(s) = 0$, since $s(x_i) = 0$, for $i = 0, 1, ..., N$ and $s^{(2q+2)}(x) = 0$, for $x_i \leq x \leq x_{i+1}$. Thus, $s^{(q+1)}(x) = 0$ in the interval $[a, b]$. Therefore, $s(x)$ is a polynomial of degree at most $q$ which has at least $N+1$, $(q \leq N)$ roots in the interval $[a, b]$. Then, $s(x) = 0$, for $x \in [a, b]$ and $a_{-m} = a_{-m+1} = ... = a_{N-1} = 0$. This means that $s(x)$ is a unique interpolating spline in $S_m(\Delta, m-1)$.
The following theorem holds:

**Theorem 3.2** *There exists a unique solution $s \in S_m(\Delta, m-1)$ of the variational problem.*

**Proof.** Let $s \in S_m(\Delta, m-1)$ be the unique interpolating spline which satisfies the conditions (3.3). Then, we note that

$$\begin{aligned} \int_a^b [g^{(q+1)}]^2 dx &= \int_a^b [s^{(q+1)}]^2 dx + \int_a^b [g^{(q+1)} - s^{(q+1)}(x)]^2 dx \\ &+2\int_a^b s^{(q+1)}(x)(g^{(q+1)}(x) - s^{(q+1)}(x))dx. \end{aligned}$$

Integrating by parts under the conditions (3.3), we find

$$\begin{aligned} &\int_a^b s^{(q+1)}(x)[g^{(q+1)}(x) - s^{(q+1)}(x)]dx \\ &= \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} s^{(q+1)}(x)[g^{(q+1)}(x) - s^{(q+1)}(x)]dx = \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &= (-1)^q \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} s^{(2q+1)}(x)[g'(x) - s'(x)]dx \\ &= (-1)^{q+1} \sum_{i=0}^{N-1} \int_{x_i}^{x_{i+1}} s^{(2q+2)}(x)[g(x) - s(x)]dx. \end{aligned}$$

Recalling the equality

$$s^{(2k+2)}(x) = 0 \quad \text{for} \quad x_i \leq x \leq x_{i+1}, \quad i = 0, 1, ..., N,$$

we find

$$\int_a^b s^{(q+1)}(x)(g^{(q+1)}(x) - s^{(q+1)}(x))dx = 0.$$

and

$$F(g) = \int_a^b [g^{(q+1)}(x)]^2 = \int_a^b [s^{(q+1)}(x)]^2 dx - \int_a^b [g^{(q+1)}(x) - s^{(q+1)}(x)]^2 dx$$

Hence, the functional $F(g)$ attains minimum at $g(x) = s(x),\ x \in [a, b]$.
In order to prove that there is a unique spline which minimizes the functional $F(g)$, let us assume for contrary that there are at least two such splines $s_1(x)$ and $s_2(x)$. Then, the difference $s(x) = s_1(x) - s_2(x)$ is the spline which satisfies homogeneous interpolation conditions. Therefore, $s(x) \equiv 0, \quad x \in [a, b]$.

**Example 3.3** *Let us consider the space $S_3(\Delta, 2)$ of cubic splines. Then, the interpolation conditions (3.3) take the following form:*

$$g(x_i) = f(x_i), \quad i = 0, 1, ..., N, \quad g^{''}(a) = g^{''}(b) = 0. \tag{3.4}$$

*By theorems (4.2) and (4.3), there exists a unique cubic spline $s \in S_3(\Delta, 2)$ at which the functional*

$$F(g) = \int_a^b [g^{''}(x)]^2 dx, \quad g \in C^{(2)}[a, b],$$

*attains minimum under the interpolation conditions (3.4), i.e.*

$$F(s) \leq \int_a^b [g^{''}(x)]^2 dx, \quad for\ all \quad g \in C^{(2)}[a, b].$$

## 3.3 Examples of Polynomial Splines

**Space $S_1(\Delta, 0)$ of Piecewise Linear Splines.** Elements of the space $S_1(\Delta, 0)$ are piecewise linear splines of the following form:

$$s(x) = a_0\psi_0(x) + a_1\psi_1(x) + \cdots + a_N\psi_N(x),$$

where the basis splines

$$\psi_i(x) = \begin{cases} \dfrac{x - x_{i-1}}{x_i - x_{i-1}} & if \quad x_{i-1} \leq x < x_i \\ \dfrac{x_{i+1} - x}{x_{i+1} - x_i} & if \quad x_i \leq x < x_{i+1} \\ 0 & if x < x_{i-1} \ \text{or} \ x \geq x_{i+1} \end{cases} \tag{3.5}$$

For the uniform distribution of the points $x_i = ih, \quad i = 0, \pm 1, \pm 2 \ldots,$ the piecewise linear splines $\psi_i(x)$ are given by the following formulae:

$$\psi_i(x) = \frac{1}{h} \begin{cases} x - x_{i-1} & if \quad x_{i-1} \leq x < x_i \\ x_{i+1} - x & if \quad x_i \leq x < x_{i+1} \\ 0 & if x < x_{i-1} \;\; or \;\; x \geq x_{i+1} \end{cases} \tag{3.6}$$

Let us note that there are $N+1$ piecewise linear splines not identically equal to zero on the interval $[a, b]$. Thus, the space of piecewise linear splines

$$S_1(\Delta, 0) = span\{\psi_0, \psi_1, \ldots, \psi_N\}$$

has dimension $N+1$.
Now, we observe that every piecewise linear spline $s(x)$ can be written as the following linear combination:

$$s(x) = a_0\psi(x) + a_1\psi_1(x) + \cdots + a_N\psi_N(x).$$

In particular, we find the Lagrange's interpolating piecewise linear spline

$$s(x) = f(x_0)\psi_0(x) + f(x_1)\psi_1(x) + f(x_2)\psi_2(x) + \cdots + f(x_N)\psi_N(x),$$

to a given function $f(x)$ which satisfies the following conditions of interpolation:

$$s(x_i) = f(x_i), \qquad i = 0, 1, ..., N.$$

The following theorem holds:

**Theorem 3.3** *If $f$ is a function twice continuously differentiable on the interval $[a, b]$, then the error of interpolation $f(x) - s(x)$ satisfies the inequality:*

$$\mid f(x) - s(x) \mid \leq \frac{h^2}{8} M^{(2)}, \;\; x \in [a, b].$$

**Proof.** Let $x_i \leq x \leq x_{i+1}$. Then, we have

$$f(x) - s(x) = \frac{f''(\xi_x)}{2!}(x - x_i)(x - x_{i+1}).$$

Since

$$\mid (x - x_i)(x - x_{i+1}) \mid \leq \frac{h^2}{4},$$

we get

$$\mid f(x) - s(x) \mid \leq \frac{h^2}{8} M^{(2)},$$

where

$$M^{(2)} = \max_{a \leq x \leq b} \mid f''(x) \mid .$$

**Example 3.4** *Approximate the function*

$$f(x) = \sqrt{1+x}, \quad 0 \le x \le 2$$

*by a piecewise linear spline with accuracy* $\epsilon = 0,01$.

**Solution.** We shall start by determining $h$. The error of interpolation of a function $f(x)$ by a piecewise linear spline $s(x)$ satisfies the following inequality:

$$| f(x) - s(x) | \le \frac{h^2}{8} M^{(2)}, \quad a \le x \le b.$$

Since

$$M^{(2)} = \max_{a \le x \le b} | f''(x) | = \max_{a \le x \le b} \frac{1}{4\sqrt{(1+x)^3}} = \frac{1}{4},$$

we may choose $h$ such that

$$\frac{h^2}{8} M^{(2)} = \frac{h^2}{32} \le \epsilon = 0.01.$$

Hence, we find $h = 0.5$.
For $h = 0.5, \; N = 4, \; x_0 = 0, \; x_1 = 0.5, \; x_2 = 1; x_3 = 1.5,$ and $x_4 = 2$, the piecewise linear spline is given by the following formula

$$\begin{aligned} s(x) = \; & (2x(\sqrt{1.5} - 1) + 1)\theta_0(x) \\ & +(2x(\sqrt{2} - \sqrt{1.5}) + 2\sqrt{1.5} - \sqrt{2})\theta_1(x) \\ & +(2x(\sqrt{2.5} - \sqrt{2}) + 3\sqrt{2} - 2\sqrt{2.5})\theta_2(x) \\ & +(2x(\sqrt{3} - \sqrt{2.5}) + 4\sqrt{2.5} - 3\sqrt{3})\theta_3 = \end{aligned}$$

$$= \begin{cases} 0.4495x + 1 & \text{if} \quad 0 \le x \le 0.5, \\ 0.3789x + 1.0353 & \text{if} \quad 0.5 \le x \le 1, \\ 0.3339x + 1.0804 & \text{if} \quad 1 \le x \le 1.5, \\ 0.3018x + 1.1284 & \text{if} \quad 1.5 \le x \le 2. \end{cases}$$

Now, let us solve this example using the following *Mathematica* module

**Program 3.1** *Mathematica module that finds linear spline for a given data table.*

---

```
linearSpline[f_,a_,b_,n_,xstep_]:=Module[{h,sol},
   h=(b-a)/n;

   onex[x_]:=Module[{xr,r},
     xr=Table[a+r*h,{r,0,n+1}];
     k=Floor[(x-a)/h+1];
     N[f[xr[[k]]]+(f[xr[[k+1]]]-f[xr[[k]]])*(x-xr[[k]])/h]
   ];

    Print["   Linear spline approximating f(x) "];
    Print[" --------------------------------------------"];
    linear=Table[{N[t],onex[t],N[f[t]]}, {t,a,b,xstep}];
    TableForm[PrependTo[linear,{" x "," linear ",f[x]}]]
   ]
```

---

By executing the following instructions

```
f[x_]:=Sqrt[1+x];
linearSpline[f,0,2,4,0.2];
```

we obtain the following table:

```
    linearSpline[f,0,2,4,0.2]
    Linear spline approximating f(x)
    ----------------------------------------
   Out[3]/TableForm=
   x       linear     Sqrt[1+x]

   0       1.         1.
   0.2     1.0899     1.09545
   0.4     1.1798     1.18322
   0.6     1.26264    1.26491
   0.8     1.33843    1.34154
   1.      1.41421    1.41421
   1.2     1.48098    1.48324
   1.4     1.54775    1.54919
   1.6     1.61132    1.61245
   1.8     1.67169    1.67322
   2.      1.73205    1.73205
```

**Space $S_3(\Delta, 2)$ of Cubic Splines.** As we know, a function $f$ can be approximated by a piecewise linear spline with the accuracy $O(h^2)$. More accurate approximation of a smooth function can be found in the space $S_3(\Delta, 2)$ of cubic splines. To determine a base of the space $S_3(\Delta, 2)$, we start with the auxiliary function

$$K_3(x_i, t) = \Delta^4 (x_i - t)_+^3, \quad x_i = ih, \quad i = 0. \pm 1 \pm 2, \ldots,$$

where

$$(x_i - t)_+^3 = \begin{cases} (x_i - t)^3 & \text{if } x \le t, \\ 0 & \text{if } x > t. \end{cases}$$

We can consider the following cubic splines as the basis of the space $S_3(\Delta, 2)$ when the points $x_i,\ i = 0, \pm, 1, \pm 2, ...,$ are uniformly distributed.

$$B_i(x) = \frac{1}{h^3} K_3(x_{i-2}, x), \quad i = -1, 0, 1, 2, \ldots, N+1.$$

The explicit form of the cubic splines $B_i(x),\ i = 0, \pm 1, \pm 2, ...,$ is given below:

$$B_i(x) = \frac{1}{h^3} \begin{cases} 0 \\ \text{if } x \le x_{i-2} \\ (x - x_{i-2})^3 \\ \text{if } x_{i-2} \le x \le x_{i-1}, \\ h^3 + 3h^2(x - x_{i-1}) + 3h(x - x_{i-1})^2 - 3(x - x_{i-1})^3 \\ \text{if } x_{i-1} \le x \le x_i, \\ h^3 + 3h^2(x_{i+1} - x) + 3h(x_{i+1} - x)^2 - 3(x_{i+1} - x)^3 \\ \text{if } x_i \le x \le x_{i+1}, \\ (x_{i+2} - x)^3 \\ \text{if } x_{i+1} \le x \le x_{i+2}, \\ 0 \\ \text{if } x \ge x_{i+2}, \end{cases} \tag{3.7}$$

for $i = 0, \pm 1, \pm 2, \ldots. \quad x \in (-\infty, \infty)$.
Now, let us note that the splines $B_i(x) = 0$ for $x \le x_{i-2}$ or $x \ge x_{i+2}$, so that the only non-zero cubic splines in the interval $[a, b]$ are $B_i(x)$ for $i = -1, 0, 1, \ldots, N+1, N+2,$ . Therefore, any cubic spline $s(x)$ can be represented on the interval $[a, b]$ as a linear combination of cubic splines $B_i(x),\ i = -1, 0, 1, ..., N+1$ i.e.,

$$s(x) = a_{-1}B_{-1}(x) + a_0B_0(x) + \cdots + a_NB_N(x) + a_{N+1}B_{N+1}(x)$$

for $x \in [a, b]$.

| $x$ | $x_{i-2}$ | $x_{i-1}$ | $x_i$ | $x_{i+1}$ | $x_{i+2}$ |
|---|---|---|---|---|---|
| $B_i(x)$ | 0 | 1 | 4 | 1 | 0 |
| $B_i'(x)$ | 0 | $\frac{3}{h}$ | 0 | $-\frac{3}{h}$ | 0 |
| $B_i''(x)$ | 0 | $\frac{6}{h^2}$ | $-\frac{12}{h^2}$ | $\frac{6}{h^2}$ | 0 |

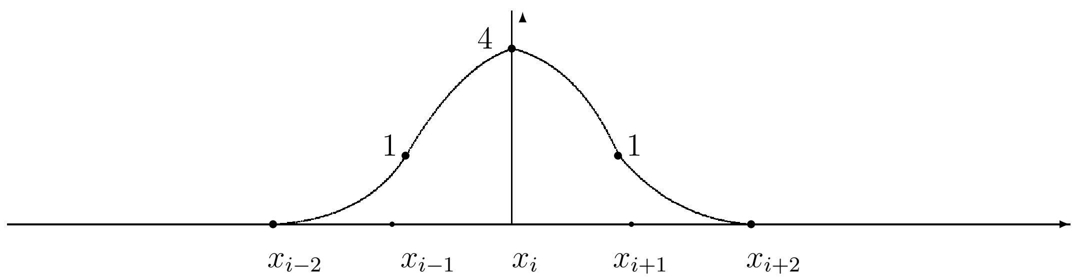



Fig. 4.3 $B_i(x)$

It can be proved that the cubic splines $B_{-1}(x), B_0(x), \ldots, B_{N+1}(x)$ create the complete basis of the space $S_3(\Delta, 2)$, so that

$$S_3(\Delta, 2) = span\{B_{-1}, B_0, \ldots, B_{N+1}\}.$$

In order to find an interpolating cubic spline $s(x)$ in the space $S_3(\Delta, 2)$, we can use the following conditions:

$$s'(x_0) = f'(x_0),$$
$$s(x_i) = f(x_i), \quad i = 0, 1, \ldots, N,$$
$$s'(x_N) = f'(x_N),$$

where $f(x)$ is a given function on the interval $[a, b]$.
There exists only one interpolating cubic spline $s(x)$ which satisfies the above conditions. From these conditions, we get the following linear system of equations:

$$\begin{array}{l} s'(x_0) = a_{-1}B'_{-1}(x_0) + a_0B'_0(x_0) + \cdots + a_{N+1}B'_{N+1}(x_0) = f'(x_0) \\ s(x_0) = a_{-1}B_{-1}(x_0) + a_0B_0(x_0) + \cdots + a_{N+1}B_{N+1}(x_0) = f(x_0) \\ s(x_1) = a_{-1}B_{-1}(x_1) + a_0B_0(x_1) + \cdots + a_{N+1}B_{N+1}(x_1) = f(x_1) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ s(x_N) = a_{-1}B_{-1}(x_N) + a_0B_0(x_N) + \cdots + a_{N+1}B_{N+1}(x_N) = f(x_N) \\ s'(x_N) = a_{-1}B_{-1}(x_N) + a_0B_0(x_N) + \cdots + a_{N+1}B_{N+1}(x_N) = f'(x_N) \end{array} \tag{3.8}$$

where $a_{-1}, a_0, a_1, \ldots, a_{N+1}$ are unknown coefficients of the cubic spline $s(x)$. The system of equations (3.8) has a unique solution, since the matrix

$$M = \left\{ \begin{array}{llllllll} -\dfrac{3}{h} & 0 & \dfrac{3}{h} & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -\dfrac{3}{h} & 0 & \dfrac{3}{h} \end{array} \right\}$$

is non-singular.
The following theorem holds (cf. [1], [9], [18]):

**Theorem 3.4** *If $s(x)$ is interpolating cubic spline to a function $f(x)$ four times continuously differentiable on the interval $[a,b]$, then the error of interpolation satisfies the following inequality:*

$$|| f^{(r)} - s^{(r)} ||_\infty \leq \epsilon_r h^{4-r} ||f^{(4)}||_\infty, \quad r = 0,1,2,3;$$

*where*

$$\epsilon_0 = \frac{5}{384}, \;\; \epsilon_1 = \frac{1}{216}(9+\sqrt{3}) \;\; \epsilon_2 = \frac{1}{12}(3\sigma+1), \;\; \epsilon_3 = \frac{1}{2}(\sigma^2+1),$$
$$h = \max_{1\leq i\leq N}(x_i - x_{i-1}), \;\; || f ||_\infty = \inf_{\mu(\Omega)=0} \sup_{a\leq x\leq b} | f(x) |,$$

*$\mu(\Omega)$ is the measure of the set $\Omega$, and $\sigma$ is the constant which defines the normal partition of the interval $[a,b]$, and $\sigma = 1$ for the uniform partition of the interval $[a,b]$*

Now, let us determine the coefficients $a_{-1}, a_0, a_1, ..., a_{N+1}$ of the cubic spline $s(x)$. Using the table, we can reduce the system of equations (3.8). Namely, from the first equation, we find

$$a_{-1} = a_1 - \frac{h}{3}f^{'}(x_0),$$

and from the last equation, we find

$$a_{N+1} = a_{N-1} + \frac{h}{3}f^{'}(x_N).$$

Then, the other $N+1$, equations, we write in the following form:

$$\begin{aligned} &2a_0 + a_1 = \frac{1}{2}[f(x_0) + \frac{h}{3}f'(x_0)] \\ &a_{i-1} + 4a_i + a_{i+1} = f(x_i), \qquad\qquad i = 1,2,...,N-1, \\ &a_{N-1} + 2a_N = \frac{1}{2}[f(x_N) - \frac{h}{3}f'(x_N)] \end{aligned}$$

In order to solve this system of equations with tri-diagonal matrix

$$M = \left\{ \begin{array}{cccccccc} 2 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 1 & 4 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 4 & 1 & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & 4 & 1 \\ 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 2 \end{array} \right\}_{(N+1)(N+1)}$$

we apply Gauss elimination method for the vector $F = \{F_0, F_1, ..., F_N\}$ in right sides:

$$F_i = \begin{cases} \frac{1}{2}(f(x_0) + \frac{h}{3}f^{'}(x_0), & i = 0, \\ f(x_i), & i = 1, 2, ..., N-1, \\ \frac{1}{2}(f(x_N) - \frac{h}{3}f^{'}(x_N), & i = N \end{cases}$$

Then, we apply the following algorithm:

$$\begin{aligned}
&\alpha_0 = \frac{1}{2}, && \beta_0 = \frac{1}{2}F_0, \\
&for \quad i = 1, 2, ..., N-1, && \\
&\alpha_i = \frac{1}{4 - \alpha_{i-1}}, && \beta_i = \frac{F_i - \beta_{i-1}}{4 - \alpha_{i-1}}, \\
&\alpha_N = \frac{1}{2 - \alpha_{N-1}}, && \beta_N = \frac{F_N - \beta_{N-1}}{2 - \alpha_{N-1}}, \\
&a_N = \beta_N, && \\
&for \quad i = N-1, N-2, ..., 1, 0, && \\
&a_i = \beta_i - \alpha_i a_{i+1}, && \\
&a_{-1} = a_1 - \frac{h}{3}f_0^{'}, && a_{N+1} = a_{N-1} + \frac{h}{3}f_N^{'}.
\end{aligned}$$

Below, we give the `Mathematica` module based on the above algorithm which produces the table of the cubic spline for a give $f(x)$ in the interval $[a, b]$. How to invoke the module, we shall explain in the example.

**Program 3.2** *Mathematica module that finds a cubic spline in the form of a table for given data table.*

---

```
cubicSpline[f_,a_,b_,n_,tstep_]:=Module[{h,sol,sp3,onet,cub},
   h=(b-a)/n;
   sol=sp3[h];

   sp3[h_]:=Module[{xi,f1a,f1b,fx,al,be,sa,sb,s},
     xi=Table[a+i*h,{i,0,n}];
     fx=N[Map[f,xi]];

     df[x_]:=D[f[x],x];
```

```
    f1a=N[df[x]/.x->a];
    f1b=N[df[x]/.x->b];

    fx[[1]]=(fx[[1]]+h*f1a/3)/2;
    fx[[n+1]]=(fx[[n+1]]-h*f1b/3)/2;

    al[1]=1/2;
    al[i_]:=al[i]=If[i<n+1,1/(4-al[i-1]),N[1/Sqrt[3]]];

    be[1]=fx[[1]]/2;
    be[i_]:=be[i]=If[i<n+1,(fx[[i]]-be[i-1])/(4-al[i-1]),
    (fx[[n+1]]-be[n])/Sqrt[3]];

    s[n+1]=be[n+1];
    s[i_]:=s[i]=be[i]-al[i]*s[i+1];

    sol=N[Table[s[i],{i,1,n+1}]];
    sa=sol[[2]]-h*f1a/3;
    PrependTo[sol,sa];
    sb=sol[[n+1]]+h*f1b/3;
    AppendTo[sol,sb]
  ];

  onet[t_]:=Module[{ },

    k=Floor[(t-a)/h+2];
    N[ ((xr[[k+1]]-t)^3*sol[[k-1]]+
    (h^3+3*h^2*(xr[[k+1]]-t)+
    3*h*(xr[[k+1]]-t)^2-3*(xr[[k+1]]-t)^3)*sol[[k]]+
    (h^3+3*h^2*(t-xr[[k]])+
    3*h*(t-xr[[k]])^2-3*(t-xr[[k]])^3)*sol[[k+1]]+
    (t-xr[[k]])^3*sol[[k+2]])/h^3]
  ];
   Print["Cubic spline approximating f(x) "];
   Print["Coefficients of the cubic spline :",Take[sol,n+3]];
   Print["------------------------------------------------"];
   xr=Table[a+r*h,{r,-1,n+1}];
 cub=Table[{N[t],onet[t],N[f[t]]}, {t,a,b-tstep,tstep}];
 AppendTo[cub,{N[b], N[((xr[[n+2]]-b)^3*sol[[n]]+
         (h^3+3*h^2*(xr[[n+2]]-b)+
         3*h*(xr[[n+2]]-b)^2-3*(xr[[n+2]]-b)^3)*sol[[n+1]]+
         (h^3+3*h^2*(b-xr[[n+1]])+
         3*h*(b-xr[[n+1]])^2-3*(b-xr[[n+1]])^3)*sol[[n+2]]+
         (b-xr[[n+1]])^3*sol[[n+3]])/h^3],N[f[b]]}];
```

```
 TableForm[PrependTo[cub,{" x "," cubic ",f[x]}]]
]
```

---

**Example 3.5** *Find a cubic interpolating spline for the following function:*

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

*spanned on the knots:* $x_0 = 0,\ x_1 = 0.5,\ x_2 = 1,\ x_3 = 1.5,$ *and* $x_4 = 2$. *Determine an approximate value* $f(1.4)$. *Estimate the error of interpolation for the function* $f(x)$.

**Solution.** The interpolating cubic spline is

$$s(x) = a_{-1}B_{-1}(x) + a_0B_0(x) + a_1B_1(x) + \cdots + a_5B_5(x),$$

The coefficients $a_{-1},\ a_0,\ a_1,\ a_2,\ a_3,\ a_4$ and $a_5$ are determined by the following system of linear equations:

$$-6a_{-1} + 6a_1 = 1,$$

$$a_{i-1} + 4a_i + a_{i+1} = e^{ih}, \quad i = 0.1, 2, 3, 4; \quad h = 0.5,$$

$$-6a_3 + 6a_5 = e^2.$$

Solving this system of equations, we find

$$\begin{aligned} s(x) &= 0.0969067B_{-1}(x) + 0.159880B_0(x) + 0.263573B_1(x) \\ &+0.434548B_2(x) + 0.7116517B_3(x) + 1.18107B_4(x) + 1.94803B_5(x), \end{aligned}$$

and

$$s(1.4) = 4.05511, \qquad f(1.3) = e^{1.4} = 4.055199967.$$

So that the error of interpolation

$$f(1.4) - s(1.4) = 0.0000899668.$$

satisfies the estimate given in the theorem, i.e.,

$$0.0000899668 = |e^{1.4} - s(1.4)| \le \frac{5}{384}0.5^4\ e^2 = 0.006,$$

Also, we can find the cubic interpolating spline to $f(x) = e^x$ using the `Mathematica` module by executing the following instructions

```
f[x_]:=Exp[x];
cubicSpline[f,0,1,4,0.1];
```

Then, we obtain the following table:

```
cubicSpline[f,0,1,4,0.1]
Cubic spline approximating f(x)
Coefficients of the cubic spline

{0.128454, 0.16494, 0.211787, 0.271938, 0.349184,
448327, 0.575707}
-----------------------------------------------------
Out[3]/TableForm=
```

| x | cubic | Exp(x) |
|---|---|---|
| 0 | 1 | 1 |
| 0.1 | 1.10316 | 1.10517 |
| 0.2 | 1.2214 | 1.2214 |
| 0.3 | 1.34985 | 1.34986 |
| 0.4 | 1.49181 | 1.49182 |
| 0.5 | 1.64872 | 1.64872 |
| 0.6 | 1.82211 | 1.82212 |
| 0.7 | 2.01376 | 2.01375 |
| 0.8 | 2.22551 | 2.22554 |
| 0.9 | 2.45952 | 2.4596 |
| 1. | 2.71828 | 2.71828 |

One can draw the graph of the cubic spline by the following `Mathematica` instructions

```
data={{0,1},{0.1,1.10517},{0.2,1.2214}, {0.3,1.34986},
        {0.4,1.49182},{0.5,1.64872},{0.6,1.82212},{0.7, 2.01375},
    {0.8,2.22554},{0.9,2.4596}, {1.0,2.71828}};
Show[Graphics[{Line[data],Spline[data,Cubic]},PlotRange->All]];
```

**Program 3.3** *Mathematica module that finds a cubic spline in the form of a list of piecewise cubic polynomials for a given data table.*

---

```
cubicSpline[f_,a_,b_,n_]:=
   Module[{h,sol,sp3,onet,cub,xr,k,r},
   h=(b-a)/n;
   sol=sp3[h];

   sp3[h_]:=Module[{xi,f1a,f1b,fx,al,be,sa,sb,s},
```

```
  xi=Table[a+i*h,{i,0,n}];
  fx=N[Map[f,xi]];

  df[x_]:=D[f[x],x];
  f1a=N[df[x]/.x->a];
  f1b=N[df[x]/.x->b];

  fx[[1]]=(fx[[1]]+h*f1a/3)/2;
  fx[[n+1]]=(fx[[n+1]]-h*f1b/3)/2;

  al[1]=1/2;
  al[i_]:=al[i]=If[i<n+1,1/(4-al[i-1]),
          N[1/Sqrt[3]]];

  be[1]=fx[[1]]/2;
  be[i_]:=be[i]=If[i<n+1,
                (fx[[i]]-be[i-1])/(4-al[i-1]),
  (fx[[n+1]]-be[n])/Sqrt[3]];

  s[n+1]=be[n+1];
  s[i_]:=s[i]=be[i]-al[i]*s[i+1];

  sol=N[Table[s[i],{i,1,n+1}]];
  sa=sol[[2]]-h*f1a/3;
  PrependTo[sol,sa];
  sb=sol[[n+1]]+h*f1b/3;
  AppendTo[sol,sb]
];
onet[t_,k_]:=Module[{ },
  ((xr[[k+1]]-t)^3*sol[[k-1]]+
  (h^3+3*h^2*(xr[[k+1]]-t)+
  3*h*(xr[[k+1]]-t)^2-3*(xr[[k+1]]-t)^3)*sol[[k]]+
  (h^3+3*h^2*(t-xr[[k]])+
  3*h*(t-xr[[k]])^2-3*(t-xr[[k]])^3)*sol[[k+1]]+
  (t-xr[[k]])^3*sol[[k+2]])/h^3
];
xr=Table[a+r*h,{r,-1,n+1}];
cub=Table[Expand[onet[t,k]], {k,2,n+1}]
]
```

---

Solving the example 4.4, we find the cubic interpolating spline to the function $f(x) = e^x, x \in [0, 2]$, by executing the following instructions

```
f[x_]:=Exp[ x];
```

```
cubicSpline[f,0,2,4]
```

Then, we obtain the following list of piecewise cubic splines determined on the subintervals $[0, 0.5]$, $[0.5, 1]$, $[1, 1.5]$ and $[1.5, 2]$.

```
                        2          3
{1. + 1. t + 0.48864 t  + 0.21249 t ,
                                   2           3
  0.982848 + 1.10291 t + 0.282818 t  + 0.349704 t ,
                                   2           3
  0.759801 + 1.77205 t - 0.386323 t  + 0.572752 t ,
                                   2           3
  -0.541948 + 4.37555 t - 2.12199 t  + 0.958444 t }
```

## 3.4 Lagrange Interpolating Splines

In the previous section, we have introduced splines of the space $S_m(\Delta, m-1)$ giving explicit form of the basis. These splines are defined on a set of spline knots $\{x_0, x_1, ..., x_n\}$. The Lagrange's interpolating splines are determined by Lagrange's conditions of interpolation given at the Lagrange's interpolating points $\{t_0, t_1, ..., t_p\}$. In general, these two sets of points do not coincide. However, in the case of piecewise linear splines both sets of points are the same. In order to determine higher order polynomial splines, by Lagrange's interpolating conditions, one has to consider the number of conditions equal to the number of coefficients in a piecewise polynomial. For example, to determine quadratic piecewise splines $a_i x^2 + b_x + c_i,\ i = 0, 1, ..., n-1$, by Lagrange's conditions of interpolation, we have to find $3n$ coefficients for which $3n$ conditions are needed. In the following example, we shall give the Lagrange's interpolating quadratic splines spanned at the spline knots $\{x_0, x_1, ..., x_n\}$ and with two different sets of interpolating points.

**Example 3.6** *Find the piecewise quadratic polynomial spline $P_2^{(i)}(x)$ spanned on knots $x_0, x_1, ..., x_n$, which satisfies Lagrange's conditions of interpolation*

$$P_2^{(i)}(x_i) = f(x_i), \qquad i = 0, 1, ..., n,$$

*for a given function $f(x)$ in the interval $[x_0, x_n]$.*

Such a spline must be continuous and continuously differentiable in the interval $[x_0, x_n]$. First, let us consider the interpolating points the same as spline knots, that is, $t_i = x_i,\ i = 0, 1, ..., n$. By the Newton's formula, we find the piecewise quadratic polynomial

$$P_2^{(i)}(x) = f_i + \frac{f_{i+1} - f_i}{x_{i+1} - x_i}(x - x_i) + a_i(x - x_i)(x - x_{i+1}), \qquad i = 0, 1, ..., n-1,$$

which satisfies the Lagrange's conditions for any values of parameters $a_i$, and for given $f_i = f(x_i)$, $i = 0, 1, ..., n$.
The parameters $a_i$, $i = 0, 1, ..., n-1$, are determined by the following continuity conditions

$$\lim_{x \to x_i^-} \frac{dP_2^{(i-1)}(x)}{dx} = \lim_{x \to x_i^+} \frac{dP_2^{(i)}(x)}{dx}, \quad i = 1, 2, ..., n-1. \tag{3.9}$$

Hence, we obtain the equations

$$a_i(x_{i+1} - x_i) + a_{i-1}(x_i - x_{i-1}) = \frac{f_{i+1} - f_i}{x_{i+1} - x_i} - \frac{f_i - f_{i-1}}{x_i - x_{i-1}}, \quad i = 1, 2, ..., n-1.$$

In the case when $x_i = x_0 + ih$, $i = 0, 1, ..., n$, these equations lead to the following recursive formula

$$a_i = -a_{i-1} + \frac{1}{h^2}[f_{i+1} - 2f_i + f_{i-1}], \quad i = 1, 2, ..., n-1.$$

at which the initial value of the parameter $a_0$ is free, and can be arbitrarily chosen.
The following `Mathematica` module spanned on knots $x_0, x_1, ..., x_n$ tabulates the quadratic spline that approximates a given function $f(x)$

**Program 3.4** *Mathematica module that tabulates the quadratic spline approximating a given function.*

---

```
quadraticSpline1[f_,a_,b_,n_,tstep_]:=Module[{h,xi,fx,k,t},
  h=(b-a)/n;
  xi=Table[a+i*h,{i,0,n+1}];
  fx=Map[f,xi];
  al[1]=1;
  al[i_]:=al[i]=-al[i-1]+(fx[[i-1]]-2 fx[[i]]+fx[[i+1]])/h^2;

  onet[t_]:=Module[{ },
    k=Floor[(t-a)/h+1];
    N[f[a+(k-1)*h]+(f[a+(k+1)*h]-f[a+k*h])*(t-xi[[k]])/h+
    al[k]*(f[a+(k-1)*h]-2 f[a+k*h]+
    f[a+(k+1)*h])*(t-xi[[k]])*(t-xi[[k+1]])]
   ];

  Print["Quadratic1 spline approximating f(x)"];
  Print["-------------------------------------"];
  quad=Table[{N[t],onet[t],N[f[t]]},{t,a,b-tstep,tstep}];
```

```
 AppendTo[quad,{N[b],onet[b],N[f[b]]}];
 TableForm[PrependTo[quad,{"x"," quadratic1 ",f[x]}]]
]
```

How to use the `Mathematica` module, we explain by the following example:

**Example 3.7** *Let*

$$f(x) = e^x, \qquad 0 \leq x \leq 1,$$

*Tabulate the quadratic spline with step* $tstep = 0.1$ *which is spanned on knots* $x_i = i\ h, \ \ i = 0, 1, 2, 3, 4, 5,$ *and approximates* $f(x)$

When `Mathematica` is active, we enter the function

```
f[x_]:=Exp[x];
```

and call

```
quadraticSpline1[f,0,1,5,0.1];
```

where $a = 0, \ b = 1$, number of intervals $n = 5$, and $tstep = 0.1$.
Then, obtain the following table:

```
quadraticSpline1[f,0,1,5,0.1]
Out[3]/TableForm=
quadratic spline approximating f(x)
------------------------------------
x          quadratic1       E^x

0          1.               1.
0.1        1.134721         1.105171
0.2        1.221403         1.221403
0.3        1.386415         1.349859
0.4        1.491825         1.491825
0.5        1.692601         1.648721
0.6        1.822119         1.822119
0.7        2.067992         2.013753
0.8        2.225541         2.225541
0.9        2.524663         2.459601
1.         2.718282         2.718282
```

Now, let us consider $n + 2$ interpolating points

$$t_i = \begin{cases} x_0, & if \quad i = 0, \\ \frac{1}{2}(x_{i-1} + x_i), & if \quad i = 1, 2, ..., n, \\ x_n & if \quad i = n + 1. \end{cases}$$

at which $f_i = f(t_i),\ i = 0, 1, ..., n + 1$.
For the equidistance points $x_i = a+ih,\ i = 0, 1, ..., n,\ nh = b-a$, the Newton's quadratic interpolating polynomial through the points $t_i, t_{i+1}, t_{i+2}$ is

$$P_2^{(i)}(x) = f_i + \frac{1}{h}(f_{i+1} - f_i)(x - t_i) + \frac{1}{2h^2}[f_i - 2f_{i+1} + f_{i+2}](x - t_i)(x - t_{i+1}),$$
$$x_i \le x < x_{i+1},\ i = 1, 2, ..., n - 2,$$

This polynomial satisfies the Lagrange's interpolating conditions

$$P_2^{(i)}(t_i) = f_i, \quad P_2^{(i)}(t_{i+1}) = f_{i+1}, \quad P^{(i)}(t_{i+2}) = f_{i+2}, \quad i = 1, 2, ..., n - 2.$$

and the continuity conditions (3.9).
Additionally, we need to find two quadratic splines that correspond to the points $t_0, t_1$, and $t_n, t_{n+1}$ which satisfy the continuity conditions at the points $x_1$ and $x_{n-1}$. One can check that the following quadratic polynomials hold these conditions

$$P_2^{(0)}(x) = f_0 + \frac{2}{h}(f_1 - f_0)(x - t_0) + \frac{2}{3h^2}(2f_0 - 3f_1 + f_2)(x - t_0)(x - t_1),$$
$$x_0 \le x < x_1,$$

and

$$\begin{aligned} P_2^{(n-1)}(x) &= f_n + \frac{2}{h}(f_{n+1} - f_n)(x - t_n) + \\ &+ \frac{2}{3h^2}(f_{n-1} - 3f_n + 2f_{n+1})(x - t_n)(x - t_{n+1}), \quad x_{n-1} \le x \le x_n. \end{aligned}$$

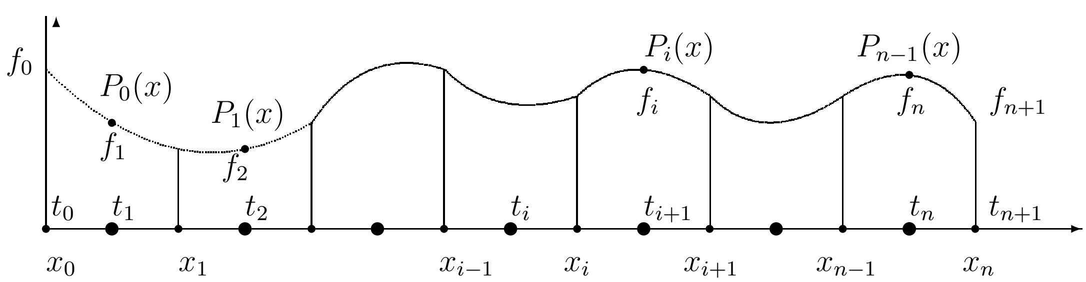


Fig. 4. Quadratic Spline

The error of interpolation is given by the formula

$$f(x) - P_2^{(i)}(x) = \frac{f'''(\xi_i)}{3!}(x - t_i)(x - t_{i+1})(x - t_{i+2}).$$

Hence, we have the following error estimate

$$|f(x) - P_2^{(i)}(x)| \leq \frac{M^{(3)}\sqrt{3}}{27}h^3, \quad x_i \leq x \leq x_{i+1}, \quad i = 0, 1, ..., n-1,$$

where $M^{(3)} = \sup_x |f^{(''')}(x)|$.
The following module in `Mathematica` tabulates the quadratic spline spanned on the Lagrange's interpolating knots $\{t_0, t_1, ..., t_n\}$ that approximates a function $f(x)$.

**Program 3.5** *Mathematica module that tabulates the Lagrange's quadratic spline approximating a given function.*

---

```
quadraticSpline2[f_,a_,b_,n_,tstep_]:=Module[{h},
   h=N[(b-a)/n];
   xi=Table[N[a+i*h],{i,0,n}];
   xt=Table[N[(xi[[i]]+xi[[i+1]])/2],{i,1,n}];
   PrependTo[xt,a];
   AppendTo[xt,b];

   p0[t_]:=Module[{ },
           f[a]+2*(f[xt[[2]]]-f[a])*(t-xt[[1]])/h+
           2*(2*f[a]-3*f[xt[[2]]]+
           f[xt[[3]]])*(t-a)*(t-xt[[2]])/(3*h^2)
          ];
   pi[t_]:=Module[{ },
              r=Floor[N[(t-a)/h]];
              f[xt[[r+1]]]+(f[xt[[r+2]]]-
              f[xt[[r+1]]])*(t-xt[[r+1]])/h+
              (f[xt[[r+1]]]-2*f[xt[[r+2]]]+f[xt[[r+3]]])*
              (t-xt[[r+1]])*(t-xt[[r+2]])/(2*h^2)
             ];
   pn[t_]:=Module[{ },
           f[xt[[n+1]]]+2*(f[xt[[n+2]]]-
           f[xt[[n+1]]])*(t-xt[[n+1]])/h+
           2*(2*f[xt[[n]]]-3*f[xt[[n+1]]]+f[xt[[n+2]]])*
   (t-xt[[n+1]])*(t-xt[[n+2]])/(3*h^2)
          ];
   onet[t_]:=Module[{ },
            Which[t<=xt[[2]],N[p0[t]],
                  t>=xt[[n+1]],N[pn[t]],True,N[pi[t]]]
           ];
```

```
    Print["   Quadratic2 spline approximating f(x) "];
    Print[" -----------------------------------------"];

 quad=Table[{N[t],onet[t],N[f[t]]}, {t,a,b,tstep}];

 TableForm[PrependTo[quad,{" x "," quadratic2 ",f[x]}]]
]
```

---

Coming back to the example 4.6, we invoke the module `quadraticSpline2` by the instructions

```
f[x_]:=Exp[x];
quadraticSpline2[f,0,1,5,0.1]
```

Then, we obtain the following table

```
      quadraticSpline2[f,0,1,5,0.1]
      Quadratic spline approximating f(x)
       -----------------------------------
   Out[2]/TableForm=

        x          quadratic2          E^x

        0          1.                  1.
        0.1        1.10517             1.10517
        0.2        1.22074             1.2214
        0.3        1.34986             1.34986
        0.4        1.49102             1.49182
        0.5        1.64872             1.64872
        0.6        1.82113             1.82212
        0.7        2.01375             2.01375
        0.8        2.22658             2.22554
        0.9        2.4596              2.4596
1.         2.71828             2.71828
```

## 3.5 Polynomial Splines of Two Variables on Rectangular Networks

Let us first consider polynomial splines associated with a rectangular network $\Delta = (\Delta_x, \Delta_y)$ which are defined an the rectangle

$$R = (x, y) \ : \ a \le x \le b, \ \ c \le y \le d,$$

where

$$\Delta x \ : \ \ a = x_0 < x_1 < \cdots < x_{N_1} = b,$$

$$\Delta y \ : \ \ c = y_0 < y_1 < \cdots < y_{N_2} = d,$$

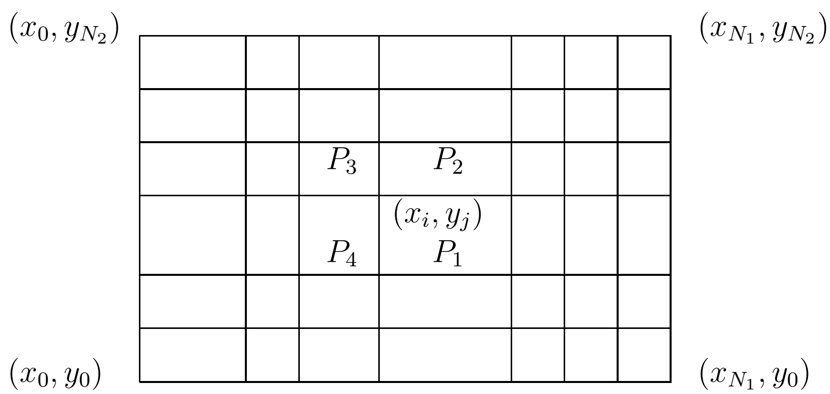



Fig. 4.5. Network $\Delta$

**Definition 3.2** *A function $s(x, y)$ is said to be a polynomial spline defined on a rectangular network if the following conditions are satisfied:*

$$(i) \quad s; \in S_{m_1}(\Delta_x, k_1) \ \ for\ any \ \ y \in [c, d],$$

$$(ii) \quad s \in S_{m_2}(\Delta_y, k_2) \ \ for\ any \ \ x \in [a, b],$$

*where $S_{m_1}(\Delta_x, k_)$ and $S_{m_2}(\Delta_y, k_2)$ are polynomial spaces of one variable (see definition 4.1) In symbols, we write $s \in S_{m_1,m_2}(\Delta, k_1, k_2)$.*

**Space $S_{11}(\Delta, 0, 0)$ of bilinear splines.** As the base of the space $S_{11}(\Delta, 0, 0)$ we can consider the following products:

$$\psi_{ij}(x, y) = \psi_i(x)\psi_j(y), \ \ i = 0, 1, \ldots, N_1, \ \ j = 0, 1, \ldots, N_2, \ \ (x, y) \in R,$$

where

$$\psi_0(x), \ \psi_1(x), \ \ldots, \psi_{N1}(x)$$

is the basis of the space $S_1(\Delta_x, 0)$,
and

$$\psi_0(y), \ \psi_1(y), \ \ldots, \psi_{N_2}(y)$$

is the basis of the space $S_1(\Delta_y, 0)$.
The explicit form of the spline $\psi_{ij}(x,y)$ is given below (cf. [16]):

$$\psi_{ij}(x,y) = \begin{cases} \dfrac{x_{i+1}-x}{x_{i+1}-x_i}\dfrac{y-y_{j-1}}{y_j-y_{j-1}} & \text{in } P_1, \\ \dfrac{x_{i+1}-x}{x_{i+1}-x_i}\dfrac{y_{j+1}-y}{y_{j+1}-y_j} & \text{in } P_2, \\ \dfrac{x-x_{i-1}}{x_i-x_{i-1}}\dfrac{y_{j+1}-y}{y_{j+1}-y_j} & \text{in } P_3, \\ \dfrac{x-x_{i-1}}{x_i-x_{i-1}}\dfrac{y-y_{j-1}}{y_j-y_{j-1}} & \text{in } P_4, \\ 0 & \text{in } R-(P_1 \cup P_2 \cup P_3 \cup P_4). \end{cases}$$

for $i = 0, 1, \ldots, N_1, \;\; j = 0, 1, \ldots, N_2$.
Thus, we can write any piecewise linear spline given on a rectangular network $\Delta$ in the form of the following linear combination:

$$s(x,y) = \sum_{i=0}^{N_1}\sum_{j=0}^{N_2} a_{ij}\psi_{ij}(x,y).$$

**Interpolation in the space** $S_{11}(\Delta, 0, 0)$. Let $f(x,y)$ be a given function twice continuously differentiable in the closed rectangle $R$. The interpolating spline $s \in S_{11}(x,y)$ which satisfies the conditions:

$$s(x_i, y_j) = f(x_i, y_j); \quad i = 0, 1, \ldots, N_1; \quad j = 0, 1, \ldots, N_2,$$

has the following form:

$$s(x,y) = \sum_{i=0}^{N_1}\sum_{j=0}^{N_2} f(x_i, y_j)\psi_{ij}(x,y).$$

By the theorem on interpolation (cf. [16]), the error $s(x,y) - f(x,y)$ satisfies the following inequality:

$$|| \, s^{(r)} - f^{(r)} \, || \leq \alpha_r h^{2-r}, \;\; r = 0, 1,$$

where $h = \max\{h_x, h_y\}, \;\; h_x = \max\{x_{i+1} - x_i\}, \;\; h_y = \max\{y_{j+1} - y_j\}; \;\; \alpha_r, \;\; r = 0, 1,$ are constants independent of $h$.
**Space** $S_{33}(\Delta, 2, 2)$ **of bicubic splines.** The space $S_{33}(\Delta, 2, 2)$ is determined by the following linearly independent set of cubic splines:

$$B_{ij}(x,y) = B_i(x)B_j(y),$$

$$i = -1, 0, 1, \ldots, N_1 + 1, \;\; j = -1, 0, 1, \ldots, N_2 + 1, \;\; (x,y) \in R,$$

where $B_i(x)$, $i = -1, 0, 1, \ldots, N_1 + 1$, and $B_j(y)$, $j = -1, 0, 1, \ldots, N_2 + 1$ are the corresponding basis of the spaces $S_3(\Delta_x, 2)$ and $S_3(\Delta_y, 2)$. Therefore, any bicubic spline given on the rectangular network $\Delta$ has the following form:

$$s(x, y) = \sum_{i=-1}^{N_1+1} \sum_{j=-1}^{N_2+1} a_{ij} B_{ij}(x, y).$$

**Interpolation in the space $S_{33}(\Delta, 2, 2)$.** Let $f(x, y)$ be a function four times continuously differentiable in the closed rectangle $R$. Then, the interpolating bicubic spline $s(x, y)$ to the function $f(x, y)$ is uniquely determined by the following conditions:

$$\begin{aligned}
&s(x_i, y_j) = f(x_i, y_j),\\
&i = -1, 0, 1, \ldots, N_1 + 1; \quad j = -1, 0, 1, \ldots, N_2 + 1,\\
&\frac{\partial s}{\partial x}(x_i, y_j) = \frac{\partial f}{\partial x}(x_i, y_j), \quad i = 0, N_1, \quad \text{when} \quad j = 0, 1, \ldots, N_2,\\
&\frac{\partial^2 s}{\partial x \partial y}(x_i, y_j) = \frac{\partial^2 f}{\partial x \partial y}(x_i, y_j), \quad i = 0, N - 1, \quad \text{and} \quad j = 0, N_2.
\end{aligned}$$

By the theorem on interpolation (cf. [16]), the error $s(x, y) - f(x, y)$ satisfies the following inequalities:

$$\begin{aligned}
&\| s - f \|_\infty \leq \beta_) h^4,\\
&\| \frac{\partial(s - f)}{\partial x} \|_\infty \leq \beta_0 h^3, \qquad \| \frac{\partial(s - f)}{\partial y} \| \leq \beta_1 h^3,\\
&\| \frac{\partial^2(s - f)}{\partial x^2} \|_\infty \leq \beta_2 h^2, \qquad \| \frac{\partial^2(s - f)}{\partial y^2} \| \leq \beta_3 h^2,\\
&\| \frac{\partial^2(s - f)}{\partial x \partial y} \|_\infty \leq \beta_4 h^2,
\end{aligned}$$

where $\beta_i$, $i = 0, 1, 2, 3, 4$ are constants independent of h.

## 3.6 Space $\Pi_1$ of Piecewise Linear Polynomial Splines on Triangular Networks

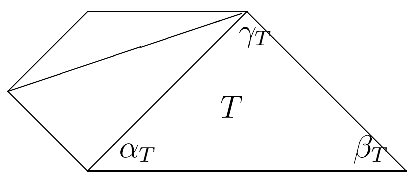
The piecewise linear splines on a triangular network have a simple structure and they are applicable to domains of arbitrary shape (cf. [23]). Let us consider the following triangularization of a bounded domain $\Omega$:

$$\Delta = \{T_0, T_1, \ldots, T_n\},$$

where

(i) $\overline{\Omega} = \overline{T_0} \cup \overline{T_1} \cup \cdots \cup \overline{T_n}$
(ii) $\overline{T_i} \cap \overline{T_j} = \emptyset$ *or they have a common side or vertix if* $i \neq j$,
(iii) *there exists a positive constant* $\nu_0$ *independent of* $n$ *such that* $\nu = \min_{\alpha_T, \beta_T, \gamma_T} \geq \nu_0$, *for all* $T \in \Delta$,

here $\alpha_T$, $\beta_T$, and $\gamma_T$ are the angles of the triangle $T$.




Let

Fig. 4.6. Triangular network

$$p_k(x, y) = a_k x + b_k y + c_k, \quad (x, y) \in T_k,$$

be a piece of a linear function and let $\Pi_1(\Delta)$ be the set of all continuous piecewise linear functions of the form:

$$s(x, y) = p_k(x, y), \quad \text{for} \quad (x, y) \in T_k, \quad k = 0, 1, \ldots, n.$$

By the theorem on interpolation (cf. [23]), there exists a unique interpolating piecewise linear spline $s \in \Pi_1(\Delta)$ to a function $f(x, y)$ which satisfies the following conditions:

$$s(x_{kl}, y_{kl}) = f(x_{kl}, y_{kl}), \quad \text{for} \quad l = 0, 1, 2,$$

where $(x_{k0}, y_{k0})$, $(x_{k1}, y_{k1})$ and $(x_{k2}, y_{k2})$ are vertices of the triangle $T_k$ for $k = 0, 1, \ldots, n$.
This spline approximates $f(x, y)$ with the error $s(x, y) - f(x, y)$ which satisfies the inequality (cf. [23])

$$\| s - f \|_\infty \leq \beta h^2,$$

where $\beta > 0$ is a positive constant independent of $h$, $\quad h = \max_{T \in \Delta} \max\{l_T^0, l_T^1, l_T^2\}$, and $dsl_T^0, l_T^1$ and $l_T^2$ are sides of the triangle T.

**Space $\Pi_3(\Delta)$ of Cubic Splines on Triangular Networks.** Let us consider a piece of a cubic polynomial in the following form:

$$P_T(x, y) = \alpha_1 + \alpha_2 x + \alpha_3 y + \alpha_4 x^2 + \alpha_5 xy + \alpha_6 y^2 + \alpha_7 x^3 + \alpha_8 x^2 y + \alpha_9 xy^2 + \alpha_{10} y^3.$$

for $(x, y) \in T$.
We shall use $\Pi_3(\Delta)$ to denote the set of all piecewise continuous cubic splines such that

$$s(x, y) = p_{T_k}(x, y), \quad \text{for } (x, y) \in T_k, \quad k = 0, 1, \ldots, n.$$

In order to find an interpolating cubic spline in the space $\Pi_3(\Delta)$, we shall consider the triangle $T$ with the center at the point $Q_0$,and its vertices at points $Q_1, Q_2$ and $Q_3$.

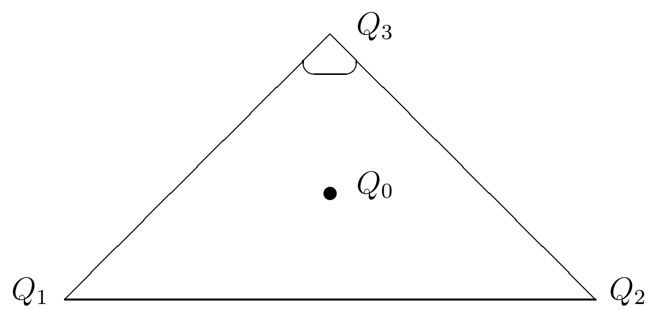



Fig. 4.7. Triangle $T$

The interpolating piecewise linear cubic spline $p_T(x, y)$ is uniquely determined in the space $\Pi_3(\Delta)$ by the following conditions (cf. [23]):

$$P_T(Q_k) = f(Q_k), \quad k = 0, 1, 2, 3,$$

$$\frac{\partial P_T(Q_k)}{\partial x} = \frac{\partial f(Q_k)}{\partial x}, \quad k = 1, 2, 3,$$

$$\frac{\partial P_T(Q_k)}{\partial y} = \frac{\partial f(Q_k)}{\partial y}, \quad k = 1, 2, 3,$$

for every $T \in \Delta$.
The following theorem holds:

**Theorem 3.5** *(cf. [23]) If $f$ is a function four times continuously differentiable in the domain $\Omega$, and $p_T \in \Pi_3(\Delta)$ is the linear piecewise interpolating cubic spline to $f$, then the error of interpolation $s - f$ satisfies the inequalities*

$$|| \ s - f \ ||_\infty \leq \frac{3M^{(4)}}{\sin\alpha} h^4,$$

$$|| \ \frac{\partial(s - f)}{\partial x} \ ||_\infty \leq \frac{5M^{(4)}}{\sin\alpha} h^3,$$

$$|| \ \frac{\partial(s - f)}{\partial y} \ ||_\infty \leq \frac{5M^{(4)}}{\sin\alpha} h^3,$$

*where* $\alpha = \min_{T\in\Delta}, \min\{\alpha_T, \beta_T, \gamma_T\} > v_0 > 0, \quad h = \max_{T\in\Delta} \max\{l_T^1, l_T^2, l_T^3\}.$

## 3.7 Exercises

**Question 3.1** *Approximate the function*

$$f(x) = ln(1 + x), \quad 0 \leq x \leq 2,$$

*by a piecewise linear spline with accuracy* $\epsilon = 0.01$. *Find an approximate value* $f(1.3)$.

**Question 3.2** *Approximate the function*

$$f(x) = \sqrt{1 + x}, \quad 0 \leq x \leq 2,$$

*and its first, second and third derivatives by a cubic spline with accuracy* $\epsilon = 0.01$.

**Question 3.3** *Give a basis of* $S_2(\Delta, 1)$ *space. Use this basis to determine a quadratic spline that approximates a function* $f(x)$.

**Question 3.4** *Write a module in Mathematica that tabulates a quadratic spline* $s \in S_2(\Delta, 1)$ *approximating the function* $f(x)$.

**Question 3.5** *Give a basis of the space* $S_5(\Delta, 4)$ *of quintic splines. Use this basis to determine a quintic spline approximating a function* $f(x)$.

**Question 3.6** *Write a module in Mathematica that tabulates a quintic spline* $s \in S_5(\Delta, 4)$ *approximating the function* $f(x)$.

**Question 3.7** *Let* $s \in S_3(\Delta, 2)$ *be a cubic spline. Following the proof of theorem 4.3, show that* $s$ *minimizes the functional*

$$F(g) = \int_a^b [g^{''}(x)]^2 dx, \quad g \in C^2[a, b],$$

*under the interpolation conditions:*

$$g(x_i) = f(x_i), \quad i = 0, 1, ..., N, \quad g^{'}(a) = s^{'}(b) = 0.$$

# Chapter 4

# Finite Element Methods

## 4.1 Introduction

There is broad literature concerning the finite element methods (cf. [3], [4], [5], [18], [26], [28], [33]). These methods are mainly based on variational principles and therefore they are associated with the Ritz and Galerkin methods. In general, Ritz or Galerkin methods may lead to a system of algebraic equations with full and ill-conditioned matrices. However, choosing proper coordinates, a system of equations with a sparse and well-conditioned matrix may be obtained. For such coordinates the Ritz or Galerkin method becomes a finite element method. In fact, finite element methods are characterized by the coordinates

$$\phi_0, \phi_1, ..., \phi_n, ...;$$

which have $\alpha-$disjoint supports, i.e., for a fixed $\alpha$

$$supp\ \phi_i(x) \cap supp\ \phi_j(x) = 0, \quad \text{when} \quad |i-j| \geq \alpha.$$

For example, a basis of polynomial splines may have $\alpha-$disjoint supports. Therefore, polynomial splines play an important role in the theory and applications of the finite element methods.

## 4.2 Finite Element Method for Elliptic Equations

Let us consider the following model of an elliptic equation:

$$Lu = f(x), \qquad x \in \Omega, \tag{4.1}$$

with the boundary value condition

$$u(x) = \phi(x), \qquad x \in \partial\Omega. \tag{4.2}$$

Equation (4.1) represents an elliptic equation if the differential operator $L$ is positive definite in the Hilbert space $H$, i.e., there exists a constant $\gamma > 0$ such that

$$(Lv, v) \geq \gamma(v, v), \qquad v \in D(L).$$

and $\overline{D(L)} = H$.
By theorem 7.2, equation (4.1) is equivalent to the following variational equation:

$$(Lu, v) = (f, v), \qquad v \in D(L). \tag{4.3}$$

In order to solve equation (4.1) by the finite element method, we define the space of finite elements

$$X_{N+1} = span\{\Phi_0, \Phi_1, ..., \Phi_N\},$$

where the basis elements $\Phi_i,\ i = 0, 1, ..., N$, have $\alpha-$disjoint supports.
We find then the approximate solution

$$u_N(x) = a_0\Phi_0(x) + a_1\Phi_1(x) + \cdots + a_N\Phi_N(x).$$

in the space $X_{N+1}$, solving the following Ritz - Galerkin system of equations:

$$(Lu_N, \Phi_k) = (f, \Phi_k), \qquad k = 0, 1, ..., N. \tag{4.4}$$

The matrix form of (4.4) is:

$$Aa = b, \tag{4.5}$$

where

$$a = [a_0, a_1, ..., a_N],\ b = [b_0, b_1, ..., b_N],\ b_k = (f, \Phi_k),\ k = 0, 1, ..., N,$$

and the Gram's matrix

$$A = \begin{bmatrix} (L\Phi_0, \Phi_0) & (L\Phi_1, \Phi_0) & (L\Phi_2, \Phi_0) & \cdots & (L\Phi_{N-1}, \Phi_0) & (L\Phi_N, \Phi_0) \\ (L\Phi_0, \Phi_1) & (L\Phi_1, \Phi_1) & (L\Phi_2, \Phi_1) & \cdots & (L\Phi_{N-1}, \Phi_1) & (L\Phi_N, \Phi_1) \\ (L\Phi_0, \Phi_2) & (L\Phi_1, \Phi_2) & (L\Phi_2, \Phi_2) & \cdots & (L\Phi_{N-1}, \Phi_2) & (L\Phi_N, \Phi_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (L\Phi_0, \Phi_N) & (L\Phi_1, \Phi_N) & (L\Phi_2, \Phi_N) & \cdots & (L\Phi_{N-1}, \Phi_{N-1}) & (L\Phi_N, \Phi_N) \end{bmatrix}$$

Because the basis elements have $\alpha-$disjoint supports, therefore $A$ is $(\alpha + 1)$-diagonal sparse matrix. Also, we note that $A$ is non-singular Gram matrix, so that, the approximate solution

$$u_N(x) = a_0^*\Phi_0(x) + a_1^*\Phi_1(x) + \cdots + a_N^*\Phi_N(x),$$

where $a_0^*, a_1^*, ..., a_N^*$, is the solution of the linear system of equations (4.5).
**Error estimate.** Let $e_N(x) = u_N(x) - u(x)$ be the global error of the method.
By theorem 7.5, the global error satisfies the inequality

$$||e_N||_L = ||u_N - u||_L \leq ||u - \eta||_L, \tag{4.6}$$

for every $\eta \in H_L$ [1]
Let $\eta(x) = s(x)$, where $s$ is the interpolating generalized polynomial to $u$.
From (4.6), we obtain the global error estimate

$$||u_n - u||_L \leq ||u - s||_L, \tag{4.7}$$

in the norm of the energetic space $H_L$.
Thus, the global error is determined by the error of interpolation, that is, if the error of interpolation

$$||u - s||_L \leq C\ h^p,$$

then the global error

$$||u_n - u||_L \leq C\ h^p. \tag{4.8}$$

### 4.2.1 Solution of Equation $-u^{''}(x) + \sigma(x)u(x) = f(x)$ in the Space of Piecewise Linear Splines.

Let us consider the equation

$$-u^{''}(x) + \sigma(x)u(x) = f(x), \qquad 0 \leq x \leq 1, \tag{4.9}$$

with the homogeneous boundary value conditions $u(0) = 0,\ \ u(1) = 0$,
where $\sigma(x) \geq 0$ and $f(x)$ are given continuous functions in interval $[0, 1]$. We shall approximate the solution $u$ by the finite element method using piecewise linear splines.
The differential operator

$$Lu \equiv -u^{''} + \sigma(x)u, \qquad u \in C_0^2(0, 1),$$

is positive definite in the space $L_2(0, 1)$, i.e., there exists $\gamma > 0$ such that

$$(Lv, v) \geq \gamma||v||^2, \qquad v \in C_0^2(0, 1).$$

By theorem 7.2, equation (4.9) is equivalent to the following variational equation

$$\int_0^1 [u^{'}(x)\eta^{'}(x) + \sigma(x)u(x)\eta(x)]dx = \int_0^1 f(x)\eta(x)dx, \quad \text{for any} \quad \eta \in W_2^{01}(0, 1), \tag{4.10}$$

We shall find the approximate solution

$$u_N(x) = a_0\psi_0(x) + a_1\psi_1(x) + \cdots + a_N\psi_N(x), \qquad 0 \leq x \leq 1.$$

in the space $S_1(\Delta, 0)$.
By the boundary conditions

$$u_N(0) = a_0 = 0, \quad \text{and} \quad u_N(1) = a_N = 0.$$

[1] $H_L = \overline{D(L)}$ is the closure of $D(L)$ in the norm $|| - ||_L$

The remaining coefficients $a_1, a_2, ..., a_{N-1}$ are determined by the variational equation

$$\int_0^1 [u_N'(x)\psi_k'(x) + \sigma(x)u_N(x)\psi_k(x)]dx = \int_0^1 f(x)\psi_k(x)dx, \qquad k = 1, 2, ..., N-1. \tag{4.11}$$

From this equation, we obtain the following system of linear equations:

$$\sum_{i=1}^{N-1} a_i \int_0^1 [\psi_i'(x)\psi_k'(x) + \sigma(x)\psi_i(x)\psi_k(x)]dx = \int_0^1 f(x)\psi_k(x)dx, \tag{4.12}$$

for $k = 1, 2, ...N-1$.
The matrix form of (4.12) is:

$$Aa = b, \tag{4.13}$$

where $a = (a_1, a_2, ..., a_{N-1})^T,\ b = (b_1, b_2, ..., b_{N-1})^T$ [2]

$$b_i = \frac{1}{h}\int_0^1 f(x)\psi_i(x)dx, \qquad i = 1, 2, ..., N-1.$$

Here, the tri-diagonal matrix $A = \{A_{ik}\},\ i, k = 1, 2, ..., N-1$ has the following entries:

$$A_{ik} = \frac{1}{h}\int_0^1 [\psi_i'(x)\psi_k'(x) + \sigma(x)\psi_i(x)\psi_k(x)]dx = \begin{cases} \dfrac{2}{h^2} + \sigma_{ii} & if \quad i = k, \\ -\dfrac{1}{h^2} + \sigma_{ik} & if \quad |i-k| = 1, \\ 0 & if \quad |i-k| \geq 2. \end{cases}$$

where $\sigma_{ik} = \dfrac{1}{h}\displaystyle\int_0^1 \sigma(x)\psi_i(x)\psi_k(x)dx$.
The matrix $A$ is diagonally dominant and positive definite. Therefore, $A$ is non-singular matrix and the unique solution $a_1^*, a_2^*, ..., a_{N-1}^*$ of (4.13) determines the approximate solution

$$u_N(x) = a_1^*\psi_1(x) + a_2^*\psi_2(x) + \cdots + a_{N-1}^*\psi_{N-1}(x).$$

In order to solve the equation (4.9) with boundary conditions, we can use the following *Mathematica* module:

```
Options[uxxTriDiag]={bound->{0,0}};

   uxxTriDiag[f_,sigm_,a_,b_,n_,opts___]:=
      Module[{a11,b11,c11,f11,i,h,t,psi1,psi2, boundary},
        h=N[(b-a)/n];
```

[2] $A^T$ denotes transposed matrix $A$

```
boundary=bound/.{opts}/.Options[uxxTriDiag];

      f0[x_]:=f[x]-sigm[x]*((boundary[[2]]-boundary[[1]])
              (x-a)/(b-a)+boundary[[1]]);

      psi1[t_,i_]:=(t-(i-1)*h)/h;
      psi2[t_,i_]:=((i+1)*h-t)/h;

      a11=Table[2/h^2+Integrate[sigm[t]*psi1[t,i]^2,
               {t,(i-1)*h,i*h}]/h+
      Integrate[sigm[t]*psi2[t,i]^2,{t,i*h,(i+1)*h}]/h,{i,1,n-1}];
      b11=Table[-1/h^2+Integrate[sigm[t]*psi2[t,i]*psi1[t,i+1],
               {t,i*h,(i+1)*h}]/h,{i,1,n-1}];
      c11=Table[-1/h^2+Integrate[sigm[t]*psi2[t,i]*
      psi1[t,i+1],{t,i*h,(i+1)*h}]/h,{i,1,n-1}];
      f11=Table[Integrate[f0[t]*psi1[t,i]/h,{t,(i-1)*h,i*h}]+

  solveTri[aa_,bb_,cc_,ff_]:=Module[{al,be,m,x},
     m=Length[aa];
     al[1]=bb[[1]]/aa[[1]];
     al[i_]:=al[i]=bb[[i]]/(aa[[i]]-al[i-1]*cc[[i]]);
     be[1]=ff[[1]]/aa[[1]];
     be[i_]:=be[i]=(ff[[i]]-be[i-1]*cc[[i]])/
             (aa[[i]]-al[i-1]*cc[[i]]);
     x[m]=be[m];
     x[i_]:=x[i]=be[i]-al[i]*x[i+1];
     Table[x[i],{i,1,m}]
     ];
      Print[" Approximate  Solution"];
      sol=solveTri[a11,b11,c11,f11]+
      Table[(boundary[[2]]-boundary[[1]])(i*h-a)/(b-a)+boundary[[1]],
      {i,1,n-1}];

     PrependTo[sol,boundary[[1]]]; AppendTo[sol,boundary[[2]]];
     Chop[sol]
   ]
```

We invoke the module by the command

```
uxxTriDiag[f,sigm,a,b,n]
```

when homogeneous boundary conditions are present, and by the commad

```
uxxTriDiag[f,sigm,a,b,n,bound ->homogeneous]
```

when non- homogeneous conditions are given.

**Example 4.1** *Let us solve with the module the following boundary value problem*

$$-\frac{d^2u(x)}{dx} + 2u(x) = \sin \pi x, \qquad 0 \leq x \leq 1,$$

*with the homogeneous boundary value conditions* $u(0) = u(1) = 0$, *using* $n = 10$ *poitnts in the interval* $[a, b]$.

We input data

```
n=10; a=0; b=1;
sigm[x_]:=2;
f[x_]:=Sin[Pi*x];
```

and execute the command

```
uxxTriDiag[f,sigm,a,b,n]
```

Then, we obtain the following list of the coefficients $\{a_0, a_1, ..., a_{10}\}$ as the output:

```
{0,0.0260703,0.0495886,0.068253,0.0802360,
   0.0843651,0.080236,0.068253,0.049589,0.026070,0}
```

Thus, the approximate solution is:

$$\begin{aligned} u_{10}(x) = \; & 0.0260703 * \psi_1(x) + 0.0495886 * \psi_2(x) + 0.068253 * \psi_3(x) + \\ & +0.0802360)\psi_4(x) + 0.0843651 * \psi_5(x) + 0.080236 * \psi_6(x) \\ & +0.068253 * \psi_7(x) + 0.049589 * \psi_8(x) + 0.026070 * \psi_9(x) \end{aligned}$$

Also, we shall solved the equatiion with the nono-hpmogeneous boundary valus conditions $u(0) = 1$ and $u(1) = 2$, by executing the following command:

```
uxxTriDiag[f,sigm,a,b,n,bound->{1,2}]
```

Then, we obtain the list of the coefficients $\{a_0, a_1, ..., a_{10}\}$ as the output:

```
{1,1.072357,1.152386,1.238119,1.328216,1.422169,
   1.520433,1.624478,1.736759,1.860596,2}
```

In this case, the approximate solution is the linear combination of eleven piecewise linear splins

$$\begin{aligned} u_{10}(x) = \; & \psi_0(x) + 1.072357 * \psi_1(x) + 1.152386 * \psi_2(x) + 1.238119 * \psi_3(x) + \\ & +1.328216\psi_4(x) + 1.422169 * \psi_5(x) + 1.520433 * \psi_6(x) \\ & +1.624478 * \psi_7(x) + 1.736759 * \psi_8(x) + 1.860596 * \psi_9(x) + 2 * \psi_{10}(x). \end{aligned}$$

**Error estimate.** By theorem (2.4), the global error $e_N = u - u_N$ satisfies the inequality

$$||e_N||_L = ||u - u_N||_L \leq ||\eta - u||_L,$$

for any $\eta \in S_1(\Delta, 0)$, where

$$||e_N||_L^2 = \int_0^1 [(e_N^{'})^2 + \sigma(x) e_N^2] dx.$$

Let $\eta(x) = s(x)$, where $s(x)$ is the piecewise linear spline interpolating the theoretical solution $u(x)$. By theorem 8.4

$$||s - u||_L \leq C\ h||u^{''}||, \qquad ||u^{''}|| = \sqrt{\int_0^1 |u^{''}|^2 dx} \tag{4.14}$$

Hence, the global error satisfies the inequality

$$||e_N||_L = ||u - u_N||_L \leq C\ h||u^{''}||. \tag{4.15}$$

where $C$ is a generic constant.
In order to obtain an estimate of $e_N$ in the norm of the space $L_2(0,1)$, we apply the Nitsche procedure (cf. [27]). Let $z \in W_2^{01}(0,1)$ be the solution of the equation

$$(Lz, \eta) = (e_N, \eta), \qquad \eta \in W_2^{01}(0,1). \tag{4.16}$$

Since $e_N \in H_L = W_2^{01}(0,1)$, therefore $z$ is the solution of the equation equality

$$(Lz, e_N) = (e_N, e_N) = ||e_N||^2, \tag{4.17}$$

where $|| - ||$ is the norm in $L_2(0,1)$.
Let $z_N \in S_1(\Delta, 0)$ be the piecewise linear spline interpolating to $z$.
Then, $z_N$ is orthogonal to $e_N$, i.e.,

$$(z_N, e_N)_L = \int_0^1 [z_N^{'} e_N^{'} + \sigma\ z_N e_N] dx = 0.$$

Thus, we have

$$(Lz, e_N) = (L(z - z_N), e_N). \tag{4.18}$$

Hence, by Cauchy's inequality

$$|(L(z - z_N), e_N)| \leq ||z - z_N||_L\ ||e_N||_L, \tag{4.19}$$

and by inequality (4.14)

$$||z - z_N||_L \leq C\ h\ ||z^{''}||. \tag{4.20}$$

Since $z$ is the weak solution of equation (4.16), we have

$$||z^{''}|| = ||e_N - \sigma\ z|| \leq ||e_N|| + ||\sigma||\ ||z||.$$

Hence, by the inequality

$$\gamma||z||^2 \leq (Lz, z) = (e_N, z) \leq ||e_N||\ ||z||,$$

we get

$$||z|| \leq \frac{1}{\gamma}||e_N||,$$

and

$$||z^{''}|| \leq C_0\ ||e_N||, \tag{4.21}$$

where $C_0 = 1 + \frac{1}{\gamma}||\sigma||$.
By inequalities (4.20) and (4.21)

$$||z - z_N||_L \leq C\ h\ ||e_N||. \tag{4.22}$$

From (4.17), (4.18) and (4.19), we get

$$||e_N||^2 \leq ||z - z_N||_L\ ||e_N||_L. \tag{4.23}$$

Finally, from (4.15), (4.22) and (4.23), we obtain the global error estimate in the norm of the space $L_2(0,1)$

$$||e_N|| = ||u - u_N|| \leq C\ h^2||u^{''}||, \tag{4.24}$$

where $C$ is a generic constant independent of $h$.

### 4.2.2 Solution of Equation $-u^{''}(x) + \sigma(x)u(x) = f(x)$ in the Space of Cubic Splines

In the previous section, we have found the solution $u_N \in S_1(\Delta, 0)$ of equation (4.9) which approximates the theoretical solution $u \in C_0^2(0,1)$ with the global error $|u(x) - u_N(x)| = O(h^2)$. We can obtain an $O(h^4)$ accurate solution $u_N$ of (4.9) using cubic splines, provided that the fourth derivative $u^{(4)} \in L_\infty(0,1)$.
We consider the boundary value problem

$$\begin{aligned} &-u''(x) + \sigma(x)u(x) = f(x), \quad 0 \leq x \leq 1, \\ &u(0) = 0, \qquad\qquad\qquad\qquad u(1) = 0 \end{aligned} \tag{4.25}$$

We write the equivalent variational form of the boundary value problem (4.25)

$$\int_0^1 [u^{'}(x)\eta^{'}(x) + \sigma(x)u(x)\eta(x)]dx = \int_0^1 f(x)\eta(x)dx, \tag{4.26}$$

for any $\eta \in H_L = W_2^{01}(0,1)$.
We find the approximate solution $u_N(x)$ in the space of cubic splines

$$S_3(\Delta, 2) = span\{B_{-1}, B_0, ..., B_{N+1}\},$$

where $B_i(x)$ for $i=-1,0,1,,...,N+1$, is the base of the space $S_3(\Delta,2)$ given by (3.7).
Then, we find the approximate solution in the form

$$u_N(x) = a_{-1}B_{-1}(x) + a_0B_0(x) + a_1B_1(x)\ldots + a_nB_N(x) + a_{N+1}B_{N+1}$$

By the homogeneous boundary conditions

$$u_N(0) = a_{-1}B_{-1}(0) + a_0B_0(0) + a_1B_1(x)\ldots + a_nB_N(0) + a_{N+1}B(0)_{N+1} = 0,$$

$$u_N(1) = a_{-1}B_{-1}(1) + a_0B_0(1) + a_1B_1(1)\ldots + a_nB_N(1) + a_{N+1}B(1)_{N+1} = 0,$$

Because (see the table)

$$B_{-1}(0)=1, \quad B_0(0)=4, \quad B_1(0)=1, \qquad B_i(0)=0, \quad i=2,3,...,N+1$$

$$B_{N-1}(1)=1, \quad B_N1=4, \quad i=B_{N+1}(1)=1, \quad B_i(1)=0, \quad i=-1,0,...,N-2,$$

Hence, by the table

$$a_{-1}+4a_0+a_1=0, \qquad a_1=-a_{-1}-4a_0,$$

$$a_{N-1}+4a_N+a_{N+1}=0, \quad a_{N+1}=-4a_N-a_N,$$

Then, the approximate solution

$$\begin{aligned} u_N(x) &= (-a_1-4a_0)B_{-1}(x) + a_0B_0(x) + a_1B_1(x) + \ldots \\ &+ a_{N-1}[B_{N-1}(x) - B_{N+1}] + (-4a_N - a_{N+1})B_{N+1}(x) \end{aligned}$$

So that

$$\begin{aligned} u_N(x) &= a_0[B_0(x) - 4B_{-1}(x)] + a_1[B_1(x) - B_{-1}(x)]\ldots \\ &+ a_{N-1}[B_{N-1}(x) - B_{N+1}(x)] + a_N[B_N(x) - 4B_{N+1}(x)]) \end{aligned}$$

and

$$u_N(x) = a_0\overline{B}_0(x) + a_1\overline{B}_1(x) + a_2\overline{B}_2(x) + \cdots \quad + a_N\overline{B}_N(x) \qquad (4.27)$$

where

$$\begin{aligned} &\overline{B}_0(x) = B_0(x) - 4B_{-1}(x), && \overline{B}_1(x) = B_0(x) - B_{-1}(x), \\ &\overline{B}_i(x) = B_i(x), && i=2,3,...,N-2, \\ &\overline{B}_{N-1}(x) = B_{N-1}(x) - B_{N+1}(x), && \overline{B}_N(x) = B_N(x) - 4B_{N+1}(x), \end{aligned} \qquad (4.28)$$

In order to find the coefficients $a_0, a_1, ..., a_N$, we solve the variational equation

$$\int_0^1 \sum_{i=1}^{N} a_i[\overline{B}_i'(x)\overline{B}_k'(x) + \sigma(x)\overline{B}_i(x)\overline{B}_k(x)]dx = \int_0^1 f(x)\overline{B}_k(x)dx, \qquad (4.29)$$

for $k = 0, 1, ..., N$.
The equations (4.29) can be written in the matrix form

$$Aa = b, \tag{4.30}$$

where $a = (a_0, a_1, ...a_N)^T, \quad b = (b_0, b_1, ..., b_N)^T$,

$$b_i = \frac{1}{h}\int_0^1 f(x)\overline{B}_i(x)dx, \qquad i = 0, 1, ..., N,$$

and the five-diagonal matrix

$$A = \begin{bmatrix} * & * & * & & & & & \\ .* & * & * & * & & & & \\ .* & * & * & * & * & & & \\ & * & * & * & * & * & & \\ & \cdots & \cdots & \cdots & \cdots & \cdots & & \\ & & & & & & * & * \end{bmatrix}$$

has ntries $*_{ik} = \dfrac{1}{h}\displaystyle\int_0^1 [\overline{B}_i'(x)\overline{B}_k'(x) + \sigma(x)\overline{B}_i(x)\overline{B}_k(x)]dx.$
One can check that matrix $A$ is non-singular. Therefore, the system of linear equations (4.30) has unique solution $a_0^*, a_1^*, ...., a_N^*$, and then the approximate solution

$$u_N(x) = a_0^*\overline{B}_0(x) + a_1^*\overline{B}_1(x) + \cdots + a_N^*\overline{B}_N(x), \tag{4.31}$$

**Error estimate.** As we know, the operator $L$ is positive definite in the space $L_2(0,1)$, and its domain $D(L) = C_0^2(0,1)$. By theorem (2.4), the global error $e_N = u - u_N$ satisfies the inequality

$$||e_N||_L = ||u - u_N||_L \leq \; ||\eta - u||_L \quad \text{ for any } \quad \eta \in H_L = W_2^{01}(0,1). \tag{4.32}$$

Let $s \in X_{N+1}$ be the cubic interpolating spline to $u(x)$.
Then, from theorem 8.4, for $\eta = s$, and by (4.32), we get

$$||s - u||_L \leq C\; h^3||u^{(4)}||, \tag{4.33}$$

Combining (4.32) and (4.33), we get the error estimate

$$||e_N||_L = ||u - u_N||_L \leq C\; h^3\; ||u^{(4)}||, \tag{4.34}$$

This estimate in the norm of the energetic space $H_L$ leads to the following error estimate

$$||e_N|| = ||u - u_N|| \leq C\; h^4\; ||u^{(4)}||_\infty, \tag{4.35}$$

in the norm of the space $L_2(0,1)$, where $C$ is a generic constant.

### 4.2.3 Solution of Equation $-\Delta u + \sigma(x,y)u = f(x,y)$ in the Space of Bilinear Splines

Let us solve the boundary value problem

$$\begin{aligned} -\Delta u \equiv -\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} + \sigma(x,y)u = f(x,y), &\quad (x,y) \in \Omega, \\ u(x,y) = 0, &\quad (x,y) \in \partial\Omega, \end{aligned} \tag{4.36}$$

by finite element method using bilinear splines.
We assume that $\sigma(x,y) \geq 0$ and $f(x,y)$ are continuous functions given in the square $\Omega = \{(x,y) \ : \ 0 \leq x, y \leq 1\}$.
As it is known, the operator

$$Lu \equiv -\Delta u + \sigma u, \qquad u \in C_0^2(\Omega),$$

is positive definite in the space $L_2(\Omega)$, i.e., there exists a constant $\gamma > 0$ such that

$$(Lv, v) \geq \gamma ||v||^2, \qquad v \in C_0^2(\Omega).$$

Therefore, we can use Ritz's or Galerkin's method to find the approximate solution $u_N(x,y)$. Following the Galerkin's method, let us replace equation (4.36) by the variational equation

$$\int_\Omega [\frac{\partial u}{\partial x}\frac{\partial \eta}{\partial x} + \frac{\partial u}{\partial y}\frac{\partial \eta}{\partial y} + \sigma \ u \ \eta] d\Omega = \int_\Omega f \ \eta \ d\Omega, \qquad \eta \in W_2^{01}(\Omega) \tag{4.37}$$

Then, we find the approximate solution

$$u_N(x,y) = \sum_{i,j=0}^{N} a_{ij}\phi_{ij}(x,y),$$

in the space

$$X_{N\times N} = S_{11}(\Delta, 0, 0) = span\{\phi_{00}, \phi_{10}, ..., \phi_{1N}, \phi_{20}, ..., \phi_{NN}\} \subset S_{11}(\Delta, 0),$$

where

$$\phi_{ij}(x,y) = \psi_i(x)\psi_j(y), \qquad i,j = 0, 1, ..., N;$$

and $\psi_i(x),\ \psi_j(y)$ are linear splines given by formula (3.6).
Because

$$u(0,y) = a_{0j} = 0, \quad u(1,y) = a_{Nj} = 0, \qquad j = 0, 1, ..., N,$$

$$u(x,0) = a_{i0} = 0, \quad u(x,1) = a_{iN} = 0, \qquad i = 0, 1, ..., N,$$

therefore

$$u_N(x,y) = \sum_{i,j=1}^{N-1} a_{ij}\phi_{ij}(x,y), \tag{4.38}$$

where the coefficients $a_{ij},\ i,j = 1,2,...,N-1$ are determined by the Galerkin system of equations

$$\sum_{i,j=1}^{N-1} a_{ij} \int_{\Omega} [\frac{\partial \phi_{ij}}{\partial x}\frac{\partial \phi_{rs}}{\partial x} + \frac{\partial \phi_{ij}}{\partial y}\frac{\partial \phi_{rs}}{\partial y} + \sigma\ \phi_{ij}\phi_{rs}] d\Omega = \int_{\Omega} f\ \phi_{rs}\ d\Omega, \tag{4.39}$$

for $r, s = 1,2,...,N-1$.
The matrix form of (4.39) is:

$$Aa = b, \tag{4.40}$$

where

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{N-1} \end{bmatrix}, \qquad a_i = \begin{bmatrix} a_{i1} \\ a_{i2} \\ \vdots \\ a_{iN-1} \end{bmatrix}, \qquad i = 1,2,...,N-1,$$

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_{N-1} \end{bmatrix}, \qquad b_i = \begin{bmatrix} b_{i1} \\ b_{i2} \\ \vdots \\ b_{iN-1} \end{bmatrix}, \qquad i = 1,2,...,N-1,$$

and the tri-diagonal block matrix

$$A = \begin{bmatrix} A_1^1 & A_2^1 & 0 & \cdots & 0 & 0 \\ A_1^2 & A_2^2 & A_3^2 & \cdots & 0 & 0 \\ 0 & A_2^3 & A_3^3 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & A_{N-2}^{N-1} & A_{N-1}^{N-1} \end{bmatrix}.$$

The entries of matrices $A_i^i = \{A_{ij}^{is}\}$, $A_{i+1}^i = \{A_{i+1j}^{is}\}$, $A_i^{i+1} = \{A_{i+1j}^{is}\}$ are given by the following formulae:

$$A_{ij}^{rs} = \int_{\Omega} [\frac{\partial \phi_{ij}}{\partial x}\frac{\partial \phi_{rs}}{\partial x} + \frac{\partial \phi_{ij}}{\partial y}\frac{\partial \phi_{rs}}{\partial y} + \sigma\ \phi_{ij}\phi_{rs}] d\Omega. \tag{4.41}$$

Hence, we have

$$A_{ij}^{is} = \begin{cases} \frac{8}{3} + \sigma_{ij}^{is} & if \quad j = s, \\ -\frac{1}{3} + \sigma_{ij}^{is} & if \quad |j-s| = 1, \\ 0 & if \quad |j-s| \geq 2, \end{cases}$$

$$A_{i\pm 1j}^{is} = \begin{cases} -\frac{1}{3} + \sigma_{i\pm 1j}^{is} & if \quad |j-s| \leq 1, \\ 0 & if \quad |j-s| \geq 2, \end{cases}$$

$$\sigma_{i\pm 1,j}^{rs} = \int_{\Omega} \sigma(x,y)\phi_{i,i\pm 1}(x,y)\phi_{rs}(x,y) d\Omega.$$

One can check that matrix $A$ is positive definite. Therefore, the system of linear equations (4.40) has the unique solution $a^*_{ij},\ i,j=1,2,...,N-1$, and then, the approximate solution

$$u_N(x,y)=\sum_{i,j=1}^{N-1} a^*_{ij}\phi_{ij}(x,y), \qquad (x,y)\in\Omega.$$

**Error Estimate.** By theorem (2.4), the global error $e_N=u-u_N$ satisfies the inequality

$$||e_N||_L=||u-u_N||_L\le||\eta-u||_L, \tag{4.42}$$

for any $\eta\in W_2^{01}(\Omega)$, where

$$||u||_L=\sqrt{\int_\Omega[(\frac{\partial u}{\partial x})^2+(\frac{\partial u}{\partial y^2})^2+\sigma\ u^2]d\Omega.}$$

Let $s(x,y)$ be the bilinear spline interpolating the theoretical solution $u(x,y)$. Then, by the theorem on interpolation, we get

$$||s-u||_L\le C\ h\ ||\Delta u||. \tag{4.43}$$

Hence, using (4.42) with $\eta=s$, we obtain the following error estimate:

$$||e_N||_L=||u-u_N||_L\le C\ h\ ||\Delta u||, \tag{4.44}$$

where the norm

$$||\Delta u||=\sqrt{\int_\Omega[(\frac{\partial u}{\partial x})^2+(\frac{\partial u}{\partial y})^2]d\Omega,} \qquad u\in W_2^0(\Omega).$$

Applying the Nitsche procedure, we can get the following estimate of the global error $e_N$ in the norm of the space $L_2(\Omega)$:

$$||e_N||=||u-u_N||\le C\ h^2\ ||\Delta u||, \tag{4.45}$$

where $C$ is a generic constant independent of $h$.

## 4.3 Finite Element Method for Parabolic Equations.

Let us consider the following model of a parabolic equation:

$$\frac{\partial u(t,x)}{\partial t}=Lu(t,x)+f(t,x), \quad t\ge 0, \quad x\in\Omega, \tag{4.46}$$

with the initial value condition

$$u(0,x)=\phi(x), \qquad x\in\Omega, \tag{4.47}$$

and with the boundary value condition

$$u(t,x) = 0, \qquad t \geq 0, \quad x \in \partial\Omega, \tag{4.48}$$

where $\phi(x)$ and $f(t,x)$ are continuous functions given for $t \geq 0$ and $x \in \Omega = \{x = (x_1, x_2, ..., x_p) \ : \ 0 \leq x_i \leq l_i, \ \ i = 1, 2, ..., p\}$.
We assume that the operator $-L$ is positive definite in the Hilbert space $H$, i.e., there exists a positive constant $\alpha > 0$ such that

$$(-Lv, v) \geq \alpha(v, v) \quad \text{for any} \quad v \in D(L), \tag{4.49}$$

where the domain $D(L)$ of $L$ is dense everywhere in $H$, so that $\overline{D(L)} = H$.
Then, the initial boundary value problem (4.46), (4.47) and (4.48) has a unique regular solution $u(t,x)$ in $D(L)$. This solution satisfies the following inequality

$$||u||^2(t) \leq ||\phi||\ e^{-\gamma t} + \frac{1}{2\epsilon}\int_0^t e^{-\gamma(t-\tau)}||f||^2(\tau)d\tau, \qquad t \geq 0 \tag{4.50}$$

for certain $\gamma > 0$ and $\epsilon > 0$.
Indeed, multiplying (4.46) by $u$, we obtain

$$(\frac{\partial u}{\partial t}, u) = (Lu, u) + (f, u).$$

Hence, by assumption (4.49) and by the Cauchy inequality

$$\frac{1}{2}\frac{\partial ||u||^2(t)}{\partial t} \leq -\alpha\ ||u||^2(t) + ||f||(t)\ ||u||(t), \qquad t \geq 0.$$

Applying the “epsilon” inequality

$$|ab| \leq \epsilon a^2 + \frac{1}{4\epsilon}b^2, \quad \epsilon > 0,$$

we get

$$\frac{\partial ||u||^2(t)}{\partial t} \leq 2(\epsilon - \alpha)||u||^2(t) + \frac{1}{2\epsilon}||f||^2(t), \qquad t \geq 0.$$

Now, choosing $\epsilon > 0$ so small that $2(\epsilon - \alpha) = -\gamma < 0$, we obtain the following inequality:

$$\frac{\partial ||u||^2(t)}{\partial t} \leq -\gamma||u||^2(t) + \frac{1}{2\epsilon}||f||^2(t), \qquad t \geq 0,$$
$$||u||(0) = ||\phi||.$$

Hence, by integration, we get (4.50).
**Finite element approach.** In order to solve the equation (4.46) by a finite element method, we replace this equation by the variational equation

$$(\frac{\partial u}{\partial t}, v)(t) = (Lu, v)(t) + (f, v)(t), \qquad \text{for any} \qquad v \in H_L, \ \ t \geq 0, \tag{4.51}$$

where $H_L$ is the energetic space associated with the operator $L$.
We shall find the approximate solution $u_h$ of (4.51) in the following space:

$$V_N = span\{\Phi_0, \Phi_1, ..., \Phi_N\},$$

where the functions $\Phi_0, \Phi_1, ..., \Phi_N$ have $\alpha-$disjoint supports.
In order to find $u_h \in V_N$, we solve the following equation:

$$(\frac{\partial u_h}{\partial t}, v_h)(t) = (Lu_h, v_h)(t) + (f_h, v_h)(t), \qquad \text{for any} \qquad v_h \in V_N, \;\; t \geq 0. \tag{4.52}$$

Thus, we look for

$$u_h(t,x) = u_0(t)\Phi_0(x) + u_1(t)\Phi_1(x) + \cdots + u_N(t)\Phi_N(x),$$

where the coefficients $u_0(t), u_1(t), ..., u_N(t)$ are determined by the Galerkin system of equations

$$\sum_{r=0}^{N} \frac{du_r(t)}{dt}(\Phi_r, \Phi_k) = \sum_{r=0}^{N} u_r(t)(L\Phi_r, \Phi_k) + \sum_{r=0}^{N} f_r(t)(\Phi_r, \Phi_k), \qquad k = 0, 1, ..., N, \tag{4.53}$$

and

$$f_h(t,x) = f_0(t)\Phi_0(x) + f_1(t)\Phi_1(x) + \cdots + f_N(t)\Phi_N(x)$$

is the interpolating generalized polynomial to $f(t,x)$.
**Error estimate.** Let

$$e_N(t,x) = u_h(t,x) - u(t,x)$$

be the global error and let

$$z_h(t,x) = u_h(t,x) - s_h(t,x) = z_0(t)\Phi_0(x) + z_1(t)\Phi_1(x) + \cdots + z_N(t)\Phi_N(x),$$

where

$$s_h(t,x) = s_0(t)\Phi_0(x) + s_1(t)\Phi_1(x) + \cdots + s_N(t)\Phi_N(x)$$

is the interpolating generalized polynomial to $u(t,x)$ for every fixed $t \geq 0$.
Then, substituting to (4.52)

$$u_h(t,x) = z_h(t,x) + s_h(t,x),$$

we obtain the following variational equation:

$$(\frac{\partial z_h}{\partial t}, \Phi_k) = (Lz_h, \Phi_k) + R_k(t,h) \qquad k = 0, 1, ..., N, \tag{4.54}$$

with the initial value condition

$$z_h(0,x) = u_h(0,x) - s_h(0,x) = 0, \qquad x \in \Omega, \tag{4.55}$$

where $\phi_h(x) = s_h(0,x) = u_h(0,x)$ is the interpolating generalized polynomial to $\phi(x)$ and

$$R_k(t,h) = (\frac{\partial z_h}{\partial t}, \Phi_k) - (Lz_h, \Phi_k) = (Ls_h, \Phi_k) + (f_h, \Phi_k) - (\frac{ds_h}{dt}, \Phi_k). \quad (4.56)$$

Because $u$ is the solution of the variational equation, we have

$$R_k(t,h) = (\frac{\partial(u - s_h)}{\partial t}, \Phi_k) - (L(u - s_h), \Phi_k) + (f - f_h, \Phi_k).$$

Hence, by the theorem on interpolation, the term $R_k(t,h) \rightarrow 0$ when $h = \frac{1}{N} \rightarrow 0$.
Now, let us write equations (4.54) in the matrix form.
We then have

$$A\frac{dz(t)}{dt} = Bz(t) + P(t,h), \qquad z(0) = 0, \quad (4.57)$$

where

$$P(t,h) = -A\frac{ds(t)}{dt} + Bs(t) + Af, \qquad t \geq 0,$$

$$z(t) = \begin{bmatrix} z_0(t) \\ z_1(t) \\ \vdots \\ z_N(t) \end{bmatrix}. \qquad s(t) = \begin{bmatrix} s_0(t) \\ s_1(t) \\ \vdots \\ s_N(t) \end{bmatrix}, \qquad f(t) = \begin{bmatrix} f_0(t) \\ f_1(t) \\ \vdots \\ f_N(t) \end{bmatrix},$$

$$A = \begin{bmatrix} (\Phi_0,\Phi_0) & (\Phi_1,\Phi_0) & (\Phi_2,\Phi_0) & \cdots & (\Phi_{N-1},\Phi_0) & (\Phi_N,\Phi_0) \\ (\Phi_0,\Phi_1) & (\Phi_1,\Phi_1) & (\Phi_2,\Phi_1) & \cdots & (\Phi_{N-1},\Phi_1) & (\Phi_N,\Phi_1) \\ (\Phi_0,\Phi_2) & (\Phi_1,\Phi_2) & (\Phi_2,\Phi_2) & \cdots & (\Phi_{N-1},\Phi_2) & (\Phi_N,\Phi_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (\Phi_0,\Phi_N) & (\Phi_1,\Phi_N) & (\Phi_2,\Phi_N) & \cdots & (\Phi_{N-1},\Phi_{N-1}) & (\Phi_N,\Phi_N) \end{bmatrix},$$

$$B = \begin{bmatrix} (L\Phi_0,\Phi_0) & (L\Phi_1,\Phi_0) & (L\Phi_2,\Phi_0) & \cdots & (L\Phi_{N-1},\Phi_0) & (L\Phi_N,\Phi_0) \\ (L\Phi_0,\Phi_1) & (L\Phi_1,\Phi_1) & (L\Phi_2,\Phi_1) & \cdots & (L\Phi_{N-1},\Phi_1) & (L\Phi_N,\Phi_1) \\ (L\Phi_0,\Phi_2) & (L\Phi_1,\Phi_2) & (L\Phi_2,\Phi_2) & \cdots & (L\Phi_{N-1},\Phi_2) & (L\Phi_N,\Phi_2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ (L\Phi_0,\Phi_N) & (L\Phi_1,\Phi_N) & (L\Phi_2,\Phi_N) & \cdots & (L\Phi_{N-1},\Phi_{N-1}) & (L\Phi_N,\Phi_N) \end{bmatrix}$$

The Gram matrix $A$ is positive definite. Thus, $A^{-1}$ exists and we can write equation (4.57) as follows:

$$\frac{dz(t)}{dt} = A^{-1}Bz(t) + A^{-1}P(t,h), \qquad z(0) = 0, \quad (4.58)$$

Multiplying both sides of the system of equations (4.58) by $z(t) \in H, \;\; t \geq 0$, we obtain

$$\frac{1}{2}\frac{d||z||^2(t)}{dt} = (A^{-1}Bz(t), z(t)) + (A^{-1}P(t,h), z(t)), \quad t \geq 0 \quad (4.59)$$

The matrices $A^{-1}$ and $-B$ are positive definite, therefore, from (4.59) and by (4.49), we have

$$\frac{d||z||^2(t)}{dt} \leq -2\alpha_0||z||^2(t)) + 2||A^{-1}P(t,h)||\ ||z||(t), \qquad t \geq 0, \tag{4.60}$$
$$z(0) = 0.$$

for a certain $\alpha_0 > 0$.
By the "epsilon" inequality

$$||P(t,h)||\ ||z||(t) \leq \epsilon||z||^2(t) + \frac{1}{4\epsilon}||A^{-1}P(t,h)||^2.$$

We can choose $\epsilon > 0$ such that

$$2(\epsilon - \alpha_0) = -\gamma < 0.$$

Then, from inequality (4.60), we get

$$\frac{d||z||^2(t)}{dt} \leq -\gamma||z||^2(t)) + \frac{1}{2\epsilon}||A^{-1}P(t,h)||^2, \qquad t \geq 0, \tag{4.61}$$
$$||z||^2(0) = 0.$$

Integrating (4.61), we obtain the following inequality:

$$||z||^2(t) \leq \frac{1}{2\epsilon}\int_0^t ||A^{-1}P(\tau,h)||^2 e^{-\gamma(t-\tau)}d\tau, \qquad t \geq 0. \tag{4.62}$$

In the most interesting cases, the norm $||A^{-1}||$ is uniformly bounded when the dimension $N \to \infty$. Also, the term $P(t,h) \to 0$, since $R_k(t,h) \to 0$ when $h = \dfrac{l}{N} \to 0$. Therefore, by (4.62),

$$z_h(t,x) = u_h(t,x) - s_h(t,x) \ \to\ 0 \qquad \text{when} \quad h \to 0.$$

Because, the error of interpolation

$$||s_h - u||(t) \ \to\ 0 \qquad \text{when} \quad h \to\ 0,$$

and

$$||u_h - u||(t) \leq ||s_h - u||(t) + ||u_h - s_h||(t), \qquad t \geq 0,$$

the global error

$$||u_h - u||(t) \ \to\ 0 \qquad \text{when} \quad h \to\ 0, \quad t \geq 0,$$

with the rate determined by the error of interpolation.

### 4.3.1 Solution of Equation $u_t = u_{xx} + f$ in the Space of Linear Splines.

Let us consider the one-dimensional heat equation

$$\frac{\partial u(t,x)}{\partial t} = \frac{\partial^2 u(t,x)}{\partial x^2} + f(t,x), \quad t \geq 0, \quad 0 \leq x \leq 1, \tag{4.63}$$

with the initial value condition

$$u(0,x) = \phi(x), \qquad 0 \leq x \leq 1, \tag{4.64}$$

and with the boundary value conditions

$$u(t,0) = 0, \qquad u(t,1) = 0, \qquad t \geq 0. \tag{4.65}$$

where $\phi(x)$ and $f(t,x)$ are continuous functions given for $t \geq 0$ and $0 \leq x \leq 1$.
The variational equation corresponding to (4.63) is:

$$\int_0^1 \frac{\partial u(t,x)}{\partial t}\eta(x)\, dx = \int_0^1 [-\frac{\partial u(t,x)}{\partial x}\frac{\partial \eta(x)}{\partial x} + f(t,x)\eta(x)]dx, \qquad t \geq 0, \tag{4.66}$$

for any $\eta \in W_2^{01}(0,1)$.
We shall find the approximate solution $u_h(t,x)$ in the space $S_1(\Delta,0)$ in following form:

$$u_h(t,x) = u_0(t)\psi_0(x) + u_1(t)\psi_1(x) + \cdots + u_N(t)\psi(x),$$

for every $t > 0$, where $\psi_k(x), \quad k = 1,2,...,N-1,$ are linear splines given by formula (3.6).
From the boundary conditions (4.65)

$$u_h(t,0) = u_0(t) = 0 \quad \text{and} \quad u_h(t,1) = u_N(t) = 0 \quad \text{for} \quad t \geq 0.$$

The remaining coefficients $u_1(t), u_2(t), ..., u_{N-1}(t)$ satisfy the Galerkin system of equations

$$\int_0^1 \frac{\partial u_h(t,x)}{\partial t}\psi_k(x)\, dx = \int_0^1 [-\frac{\partial u_h(t,x)}{\partial x}\frac{d\psi_k(x)}{dx} + f_h(t,x)\psi_k(x)]dx, \tag{4.67}$$

for $t \geq 0, \quad k = 1,2,...,N-1$, where

$$f_h(t,x) = f(t,x_0)\psi_0(x) + f(t,x_1)\psi_1(x) + \cdots + f(t,x_{N-1})\psi_{N-1}(x) + f(t,x_N)\psi_N(x).$$

Let us rewrite (4.67) as follows:

$$\begin{aligned} &\sum_{r=1}^{N-1} \frac{du_r(t)}{dt} \int_0^1 \psi_r(x)\psi_k(x)\, dx = \\ &= \sum_{r=1}^{N-1} u_r(t) \int_0^1 -\frac{d\psi_r(x)}{dx}\frac{d\psi_k(x)}{dx}\, dx + \int_0^1 f_h(t,x)\psi_k(x)dx, \end{aligned} \tag{4.68}$$

for $t \geq 0, \quad u_k(0) = \phi(x_k), \; k = 1, 2, ..., N-1.$
Let us note that

$$\int_0^1 \psi_r(x)\psi_k(x)dx = \begin{cases} \dfrac{2h}{3} & if \quad r = k, \\ \dfrac{h}{6} & if \quad |r-k| = 1, \\ 0 & otherwise \end{cases} \tag{4.69}$$

and

$$\int_0^1 \frac{d\psi_r(x)}{dx}\frac{d\psi_k(x)}{dx}\,dx = \begin{cases} \dfrac{2}{h} & if \quad r = k, \\ -\dfrac{1}{h} & if \quad |r-k| = 1, \\ 0 & otherwise, \end{cases} \tag{4.70}$$

Thus, equations (4.68) can be written in the matrix form:

$$A\frac{du(t)}{dt} = Bu(t) + F(t), \qquad t \leq 0, \tag{4.71}$$

with the initial value condition

$$u(0) = \phi, \tag{4.72}$$

where

$$u(t) = \begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_{N-1}(t) \end{bmatrix}, \qquad \phi = \begin{bmatrix} \phi(x_1) \\ \phi(x_2) \\ \vdots \\ \phi_{N-1}(x_{N-1}) \end{bmatrix}, \qquad F(t) = \begin{bmatrix} F_1(t) \\ F_2(t) \\ \vdots \\ F_{N-1}(t) \end{bmatrix},$$

$$\begin{aligned} F_k(t) = \; & \frac{6}{h}\int_0^1 f_h(t,x)\psi_k(x)\,dx = \\ & f(t, x_{k-1}) + 4(t, x_k) + f(t, x_{k+1}), \quad k = 1, 2, ..., N-1, \end{aligned}$$

$$A = \begin{bmatrix} 4 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 4 & 1 & \cdots & 0 & 0 \\ 0 & 1 & 4 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 1 & 4 \end{bmatrix}, \quad B = -\frac{6}{h^2}\begin{bmatrix} 2 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & -1 & 2 \end{bmatrix}.$$

The matrices $A$ and $B$ are positive definite, so that, the system of equations (4.71) possesses the unique solution

$$u_1^*(t), u_2^*(t), ..., u_{N-1}^*(t).$$

which determines the approximate solution

$$u_h(t,x) = u_1^*(t)\psi_1(x) + u_2^*(t)\psi_2(x) + \cdots + u_{N-1}^*(t)\psi_{N-1}(x), \qquad t \geq 0.$$

**Error estimate.** Let $e_N(t,x) = u_h(t,x) - u(t,x)$ be the global error and let

$$z_h(t,x) = u_h(t,x) - s_h(t,x) = z_1(t)\psi_1(x) + z_2(t)\psi_2(x) + \cdots + z_{N-1}(t)\psi_{N-1}(x),$$

where

$$s_h(t,x) = s_1(t)\psi_1(x) + s_2(t)\psi_2(x) + \cdots + s_{N-1}(t)\psi_{N-1}(x)$$

is the interpolating linear spline to $u(t,x)$ for every fixed $t \geq 0$.
Then, we have

$$s_k(t) = u(t,x_k), \quad k = 1,2,...,N-1.$$

Substituting to (4.67),

$$u_h(t,x) = z_h(t,x) + s_h(t,x),$$

we obtain the following variational equation:

$$\int_0^1 \frac{\partial z_h(t,x)}{\partial t}\psi_k(x)\,dx = \int_0^1 -\frac{\partial z_h(t,x)}{\partial x}\frac{d\psi_k(x)}{dx}\,dx + R_k(t,h), \quad k = 1,2,...,N-1. \tag{4.73}$$

with the homogeneous initial value condition

$$z_h(0,x) = u_h(0,x) - s_h(0,x) = 0, \qquad 0 \leq x \leq 1. \tag{4.74}$$

Here, $\phi_N(x) = u_h(0,x) = s_h(0,x)$ is the interpolating linear spline to $\phi(x)$ and

$$\begin{aligned} R_k(t,h) &= \int_0^1 \frac{\partial z_h(t,x)}{\partial x}\frac{d\psi_k(x)}{dx} + \int_0^1 \frac{\partial z_h}{\partial x}\frac{d\psi_k}{dx}dx = \\ &= -\int_0^1 [\frac{\partial s_h(t,x)}{\partial x}\frac{d\psi_k(x)}{dx} + \frac{\partial s_h(t,x)}{\partial t}\psi_k(x) - f_h(t,x)\psi_k(x)]\,dx. \end{aligned} \tag{4.75}$$

The matrix form of (4.73) is:

$$A\frac{dz(t)}{dt} = Bz(t) + R(t,h), \qquad z(0) = 0, \tag{4.76}$$

where

$$z(t) = \begin{bmatrix} z_1(t) \\ z_2(t) \\ \vdots \\ z_{N-1}(t) \end{bmatrix}, \qquad R(t,h) = \frac{6}{h}\begin{bmatrix} R_1(t,h) \\ R_2(t,h) \\ \vdots \\ R_{N-1}(t,h) \end{bmatrix}.$$

Because the matrix $A$ is non-singular, (4.76) is equivalent to

$$\frac{dz(t)}{dt} = A^{-1}Bz(t) + A^{-1}R(t,h), \qquad z(0) = 0, \tag{4.77}$$

Multiplying both sides of the system of equations (4.77) by $z(t) \in H_h^0,\ \ t \geq 0$, we obtain the following equation;

$$\frac{1}{2}\frac{d||z||^2(t)}{dt} = (A^{-1}Bz(t), z(t)) + (A^{-1}R(t,h), z(t)), \quad t \geq 0, \tag{4.78}$$

where

$$||z||^2(t) = z_1^2(t) + z_2^2(t) + \cdots + z_{N-1}^2(t).$$

Now, let us note that

$$\begin{aligned}
&A = \frac{h^2}{6}B + 6I,\\
&(Av, v) = \frac{h^2}{6}(Bv, v) + 6(v, v) \geq 2(v, v),\\
&||A^{-1}|| \leq \frac{1}{2},\\
&(A^{-1}v, v) \geq \frac{1}{6}(v, v), \quad v \in H_h^0,\\
&(A^{-1}Bv, v) = (BA^{-\frac{1}{2}}v, A^{-\frac{1}{2}}v) \leq -\frac{48}{l^2}(A^{-1}v, v) \leq -\frac{8}{l^2}(v, v),
\end{aligned}$$

for $v \in H_h^0$. Hence, by (4.78)

$$\begin{aligned}
&\frac{d||z||^2(t)}{dt} \leq -\frac{16}{l^2}||z||^2(t)) + 2||R(t,h)||\ ||z||(t), \quad t \geq 0,\\
&z(0) = 0.
\end{aligned} \tag{4.79}$$

and by the “epsilon” inequality

$$||R(t,h)||\ ||z||(t) \leq \epsilon||z||^2(t) + \frac{1}{4\epsilon}||R(t,h)||^2.$$

We can choose $\epsilon > 0$ such that

$$2\epsilon - \frac{16}{l^2} = -\gamma < 0.$$

Then, from inequality (4.79), we obtain

$$\begin{aligned}
&\frac{d||z||^2(t)}{dt} \leq -\gamma||z||^2(t)) + \frac{1}{2\epsilon}||R(t,h)||^2, \quad t \geq 0,\\
&||z||^2(0) = 0.
\end{aligned} \tag{4.80}$$

Integrating this inequality, we get

$$||z||^2(t) \leq \frac{1}{2\epsilon}\int_0^t ||R(\tau,h)||^2 e^{-\gamma(t-\tau)}d\tau \quad t \geq 0. \tag{4.81}$$

To estimate $R(t,h)$, we shall use formulae (4.69), (4.70) and (4.75).
Then, we have

$$\begin{aligned}
\frac{6}{h}R_k(t,h) &= -\frac{6}{h}\int_0^1 \sum_{j=1}^{N-1}[s_j(t)\frac{d\psi_j}{dx}\frac{d\psi_k}{dx} + \frac{ds_j}{dt}\psi_j\psi_k - f(t,x_j)\psi_j\psi_k]dx \\
&= \frac{6}{h^2}[s_{k-1}(t) - 2s_k(t) + s_{k+1}(t)] - [\frac{ds_{k-1}}{dt} + 4\frac{ds_k}{dt} + \frac{ds_{k+1}}{dt}] \\
&+[f(t,x_{k-1}) + 4f(t,x_k) + f(t,x_{k+1})] \\
&= 6[-\frac{\partial u(t,x_k)}{dt} + \Lambda_x u(t,x_k) + f(t,x_k)] + h^2[\Lambda_t\frac{\partial u(t,x_k)}{dt} + \Lambda_x f(t,x_k)] \\
&= -\frac{h^2}{12}\frac{\partial^4 u(t,\xi_k)}{\partial x^4} + h^2[\Lambda_t\frac{\partial u(t,x_k)}{dt} + \Lambda_x f(t,x_k)] = O(h^2).
\end{aligned}$$

for a certain $\xi_k$
Hence, by inequality (4.81)

$$||z||^2(t) \le \frac{C\ h^4}{2\epsilon\ \gamma}[1-e^{-\gamma t}], \qquad t \ge 0, \tag{4.82}$$

or

$$||z||(t) \le C_0\ h^2, \qquad t \ge 0, \tag{4.83}$$

where

$$C_0 = \sqrt{\frac{C}{2\epsilon\ \gamma}}.$$

The norm

$$\begin{aligned}
||z_h||^2(t) = \int_0^1 |z_h^2(t,x)|dx = & \int_0^1 [\sum_{k=1}^{N-1} z_k(t)\psi_k(x)]^2 dx \\
& \le 2\sum_{k=1}^{N-1} z_k^2(t)\int_0^1 \psi_k^2(x)dx \le C\ h^4,
\end{aligned} \tag{4.84}$$

where $C$ is a generic constant.
Using the triangle inequality

$$||u_h - u||(t) \le ||u_h - s_h||(t) + ||s_h - u)||(t),$$

and the inequality

$$||s_h - u||(t) \le C\ h^2, \qquad t \ge 0,$$

by (4.84), we obtain the global error estimate:

$$||u_h - u||(t) \le C\ h^2, \qquad t \ge 0, \tag{4.85}$$

where C is a generic constant.

## 4.4 Boundary Conditions

So far, we have considered homogeneous Dirichlet boundary conditions for both elliptic and parabolic equations. Let us now study the non-homogeneous boundary conditions

$$A_0\frac{du(0)}{dx} + B_0u(0) = \alpha, \qquad A_1\frac{du(1)}{dx} + B_1u(1) = \beta, \tag{4.86}$$

associated with the differential equation

$$Lu \equiv -\frac{d^2u}{dx^2} + \sigma(x)u = f(x), \qquad 0 \le x \le 1, \tag{4.87}$$

where $\sigma(x) \ge 0,\ f(x)$, are given continuous functions.
In order to find finite element solution

$$u_h(x) = \psi(x) + \sum_{i=1}^{N} a_i\phi_i(x),$$

we write equation (4.87) in the variational form

$$\int_0^1 [-\frac{d^2u_h}{dx^2} + \sigma(x)u_h]\phi_i(x)dx = \int_0^1 f(x)\phi_i(x)dx, \qquad i = 1, 2, ..., N.$$

Integrating by parts

$$\int_0^1 [\frac{du_h}{dx}\frac{d\phi_i}{dx} + \sigma(x)u_h]\phi_i(x)dx = \int_0^1 f(x)\phi_i(x)dx + \frac{du_h(1)}{dx}\phi_i(1) - \frac{du_h(0)}{dx}\phi_i(0), \tag{4.88}$$

for $i = 1, 2, ..., N$.
The function $\psi(x)$ and the coordinates $\phi_i(x),\ i = 1, 2, ..., N$, must be chosen according to specific boundary conditions, that is, Dirichlet's, Neumann's or mixed conditions.
We shall investigate these conditions as special cases of (4.86).
**Dirichlet conditions.** Setting in (4.86), $A_0 = A_1 = 0$ and $B_0 = B_1 = 1$, we get the Dirichlet conditions

$$u_h(0) = \alpha, \qquad u_h(1) = \beta.$$

These non-homogeneous conditions can easily be replaced by homogeneous ones substituting to (4.88)

$$u_h(x) = \psi(x) + \sum_{i=1}^{N} a_i\phi_i(x),$$

where $\psi(x)$ is a function for which $\psi(0) = \alpha,\ \psi(1) = \beta$, and

$$v_h(x) = \sum_{i=1}^{N} a_i\phi_(x),$$

is the solution of variational equation (4.88) holding the homogeneous Dirichlet's conditions $v_h(0) = v_h(1) = 0$.
It is clear, from (4.88), that the homogeneous conditions must be imposed on the coordinates $\phi_i(x)$. Thus, we have to choose $\phi_i \in H_L$, such that $\phi_i(0) = \phi_i(1) = 0,\ i = 1, 2, ..., N$.
The conditions which must be posed on the coordinates $\phi_i,\ i = 1, 2, ..., N$, are called essential boundary conditions, and any others are referred to as natural boundary conditions.
Thus, the Dirichlet conditions are essential, since coordinates $\phi_i(x),\ i = 1, 2, ..., N$, must carry out the corresponding homogeneous conditions.
**Neumann's conditions.** Setting in (4.86) $A_0 = A_1 = 1$ and $B_0 = B_1 = 0$, we get the Neumann conditions

$$\frac{du_h(0)}{dx} = \alpha, \qquad \frac{du_h(1)}{dx} = \beta.$$

The Neumann conditions cannot be specified arbitrarily. Data in the Neumann problem must be compatible. Indeed, for $\sigma(x) \equiv 0$ and $\eta(x) \equiv 1$, the equation

$$\int_0^1 [\frac{du}{dx}\frac{d\eta}{dx}dx = \int_0^1 f(x)\eta(x)dx + \frac{du(1)}{dx}\eta(1) - \frac{du(0)}{dx}\eta(0), \quad i = 1, 2, ..., N,$$

reduces to the compatibility condition

$$\int_0^1 f(x)dx = \beta - \alpha.$$

Dealing with the Neumann problem, we note that if $u(x)$ is a solution of the Neumann problem, then $u(x) + constant$ is also a solution of the Neumann problem. Therefore, a condition must be set up on $u(x)$ to normalized a unique solution $u(x)$.
The non-homogeneous Neumann conditions can also be replaced by homogeneous ones substituting to (4.88)

$$u_h(x) = \psi(x) + \sum_{i=1}^{N} a_i\phi_i(x),$$

where $\psi(x)$ is a function that satisfies the non homogeneous Neumann conditions

$$\psi'(0) = \alpha, \quad \psi'(1) = \beta.$$

For example $\psi(x) = H_3(x)$ can be the Hermite's cubic polynomial determined by the interpolation conditions

$$H_3(0) = u_h(0), \quad H_3(1) = u_h(1), \quad H_3'(0) = \alpha, \quad H_3'(1) = \beta.$$

Then, the function

$$v_h(x) = \sum_{i=1}^{N} a_i\phi_i(x),$$

satisfies the variational equation

$$\begin{aligned}&\int_0^1 [\frac{dv_h}{dx}\frac{d\phi_i}{dx} + \sigma(x)v_h(x)\phi_i(x)]dx = \\ &= \int_0^1 \{\frac{d\psi}{dx}\frac{d\phi_i}{dx} + [f(x) - \sigma(x)\psi(x)]\phi_i(x)\}dx + \beta\phi_i(1) - \alpha\phi_i(0),\end{aligned} \tag{4.89}$$

for $i = 1, 2, ..., N$.
In the Neumann problem, we do not impose the boundary conditions on the coordinates $\phi_i \in H_L,\ i =, 1, 2, ..., N$, since they have been used directly in the equation (4.89).
**Mixed conditions.** Setting in (4.86), $A_0 = 0,\ B_0 = 1$ and $A_1 = 1,\ B_1 = 0$, we get the mixed conditions

$$u_h(0) = \alpha, \qquad \frac{du_h(1)}{dx} = \beta.$$

Similarly as for Dirichlet or Neumann problems, we find the finite element solution in the following form:

$$u_h(x) = \psi(x) + \sum_{i=1}^{N} a_i\phi_i(x),$$

where $\psi(x)$ is to be chosen in a way to satisfy the mixed conditions, i.e.,

$$\psi(0) = \alpha, \quad \psi^{'}(1) = \beta.$$

Then, the coordinates $\phi_i(x),\ i = 1, 2, ..., N$, must hold the homogeneous condition at $x = 0$, so that

$$\phi_i(0) = 0, \quad i = 1, 2, ..., N.$$

However, at $x = 1$, there is no restriction posed on $\phi_i(x)$.
In this case, the finite element solution

$$v_h(x) = \sum_{i=1}^{N} a_i\phi_i(x),$$

satisfies the variational equation

$$\int_0^1 \{\frac{dv_h}{dx}\frac{d\phi_i}{dx} + [f(x) - \sigma(x)\psi(x)]\phi_i(x)\}dx + \beta\phi_i(1),$$

for $i = 1, 2, ..., N$.
In general, non-homogeneous conditions can be replaced by homogeneous ones using the substitution

$$u_h(x) = \psi(x) + v_h(x), \quad x = (x_1, x_2, ..., x_N) \in \Omega,$$

where $\psi \in H_L$ is a function holding the non-homogeneous conditions

$$u_h(x) = \psi(x), \qquad x \in \partial\Omega.$$

and $v_h$ is the finite element solution of the modified variational equation

$$(Lv_h, \eta) = (f - L\psi, \eta), \qquad \eta \in H_L. \tag{4.90}$$

Then, the coordinates $\phi_i \in H_L,\ i = 1, 2, ..., N,$ must carry out the essential boundary conditions, while the natural conditions are used in variational equation (4.90).
Let us note that, we can easily identify the essential boundary conditions for linear differential operators of order $2s$. Then, a boundary condition is essential if it does not contain a derivative of order greater than $s-1$, otherwise it is referred to be a natural boundary condition.
For example, the Dirichlet condition $u(x, y) = g(x, y),$ at $\partial\Omega$, for the Laplace operator

$$Lu \equiv \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2},$$

is essential, since $s = 1$ and this condition does not contain a derivative ($s-1 = 0$), while the Neumann's condition $\dfrac{du}{dn} = g(x, y),$ at $\partial\Omega$, is natural one, since it contains the derivative of order $s = 1$.

## 4.5 Exercises

**Question 4.1** *Consider the following boundary value problem:*

$$-\frac{d^2u}{dx^2} + 3u = e^{-x^2}, \qquad 0 \le x \le 1,$$
$$u(0) = 0, \qquad u(1) = 0.$$

*Give an finite element approximation of the boundary problem using*

1. *(a) piecewise linear splines,*
   *(b) cubic splines.*

*Estimate the error of the method in both cases (a) and (b).*

**Question 4.2** *Use the piecewise linear splines with respect to the variable $x$ to approximate the following initial-boundary value problem:*

$$\frac{\partial u}{\partial t} = 4\frac{\partial^2 u}{\partial x^2} + \cos \pi x, \qquad 0 \le x \le 1, \quad t \ge 0,$$
$$u(0, x) = \sin \pi x, \qquad 0 \le x \le 1,$$
$$u(t, 0) = 0, \qquad u(t, 1) = 0, \quad t \ge 0,$$

*by the finite element method. Estimate the error of the method.*

## 4.6 References

[**1** ] Aubin, J. P., (1972), Approximation of Elliptic Boundary Value Problems, Wiley- Interscience

[**2** ] Berezin, I.S. & Zidkov, N.P. (1962), Numerical Methods, v.II,Moskaw

[**3** ] Hall, C.A. (1984), Numerical Analysis of Partial Differential Equations. Hemisphere Publishing Company.

[**4** ] Michlin, S.G. & Smolnicky C.L. (1963), Approximate Methods for Solving Differential and Integral Equations, Nauka, Moscow.

[**5** ] Mitchell, A.R. & Griffiths D.F. (1975), The Finite Difference Method in Partial Differential Equations, John Wiley & Sons, New York.

[**6** ] Quarteroni, A., & Valli, A., (1994), Numerical Approximation of Partial Differential Equations, Springer-Verag, New York.

[**7** ] Smith, G.D. (1985), Numerical Solution of Partial Differential Equations: Finite Difference Methods, Clarendon Press-Oxford.

[**8** ] Stoer, J. & Bulrisch R. (1980), Introduction to Numerical Analysis, Springer-Verlag, (1980).

[**9** ] Wolfram, S., (1992),Mathematica a System for Doing Mathematics by Computer, Addison-Wesley Publishing Company, New York.