Building Virtual Models in Engineering
An Introduction to Finite Elements

Fred Vermolen and Domenico Lahaye

Delft University of Technology

Faculty of Electrical Engineering, Mathematics and Computer Science

Delft Institute for Applied Mathematics

November 2011

Preface

Preface to the 2011-2012 Edition In our scientific endeavor we stand upon the shoulder of giants. Early version of these notes were written under the supervision of Frederique le Magnifique and the master himself, who in turn learned the trade from the finite element pioneers at the TU Delft. We thank Niel Budko for pointing to typos in previous versions of these notes, students at the department for their help in further developing the laboratory sessions that make integral part of this course and acknowlegde the help of all the support staff allowing this course to be continued.

Domenico Lahaye - 2011

Preface to Earlier Editions These lecture notes are used for the course Finite Elements (WI3098LR) for BSc-students of the department of aerospace engineering with a minor in applied mathematics, and for students visiting the Delft University of Technology in the framework of ATHENS. For the interested student some references are given for further reading, but note that the literature survey is far from complete.

The notes provide an introduction into the Finite Element method applied to Partial Differential Equations. The treated problems are classical. The treatment of the Finite Element Method is mathematical, but without the concept of Hilbert and Sobolev spaces, which are fundamental in a rigorous treatment of the Finite Element Method. Recently, a more detailed and comprehensive treatment appeared by van Kan et al. [2006], which is available from the Delft University Press.

Further, we recommend all students of this course to attend the lectures since the lecture notes do not aim at being complete.

We wish everybody good luck with the material and the course! Further, we would like to thank Caroline van der Lee for the beautiful typesetting of this document. Finally, we thank Fons Daalderop and Martin van Gijzen for their critical reading of these lecture notes and valuable suggestions to improve it.

- Alfonzo e Martino, grazie!

Fred Vermolen and Domenico Lahaye - 2008 - 2011

Contents

1

Introduction

In this chapter we aim at

- describing five stages that are typical in the application of a finite element method (FEM) modeling procedure for solving engineering problems;
- giving examples of how FEM is applied in the modeling of industrial furnaces and electrical energy applications;
- giving a list of so-called model problems that will be used in the remainder of this courses.

1.1 Stages in a FEM Modeling Procedure

Figure 1.1 shows five different stages in the finite element method (FEM) modeling of the mechanical deformation of a microchip caused by the Ohmic heat generated by an electrical current. The following five stages can be distinguished:

- in **Stage 1** a geometry model of the device under consideration is build. For this purpose a computer aided design (CAD) software package is typically used;
- in **Stage 2** the governing physical processes are described in term of appropriate partial differential equations supplied with initial and boundary conditions. In the modeling of electrical energy applications for instances, the Maxwell equations for the electric and magnetic field play a central role;
- in **Stage 3** the geometry is subdivided into small entities called elements. Different types of mesh generation and mesh refinement techniques exist. General references on this stage are Farrashkhalvat and Miles [2003]; Frey and George [2000]; Liseikin [2010] (mesh generation) and Ainsworth and Oden [2000] (adaptive refinement). Mesh generation techniques are being taught at the TU Delft as part of the *Elements of Computational Fluid Dynamics* course (WI4011).

- in **Stage 4** the governing partial differential equations are discretized in space and time, a time-stepping procedure is applied to the resulting system of ordinary differential equations and the emerging (non-)linear systems of equations are solved. General references on this stage are Braess [1996]; Eriksson et al. [1996]; Morton and Mayers [1994]; Quarteroni and Valli [1994]; Zienkiewics and Taylor [2000a,b] (general finite elements), Ascher and Petzold [1998] (time-stepping methods), Dennis Jr. and Schnabel [1996]; Kelley [1995] (non-linear system solvers) and Saad [1996]; Smith et al. [1996]; Trottenberg et al. [2001] (linear system solvers). Reference to the mathematical theory underlying the finite element method are for instance Brenner and Scott [2008]; Kreyszig [1989]; Atkinson and Han [2005]. Linear system solvers are being at the TU Delft in the course *Scientific Computing* (WI4210). TU Delft monographs on ordinary differential equation and on the numerical treatment of partial differential equations are Vuik et al. [2007] and van Kan et al. [2006], respectively.

- in **Stage 5** the solution of the partial differential equations and derived quantities are visualized on the mesh employed for discretization in space and time.

These five stages are typical in any FEM modeling procedure. Although the other stages are not trivial and object of active research, this course mainly focussed on **Stage 2** and **Stage 4**.

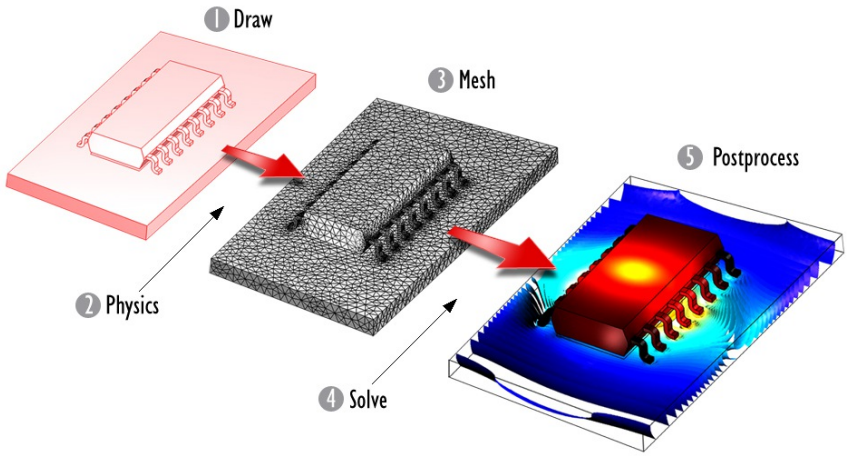
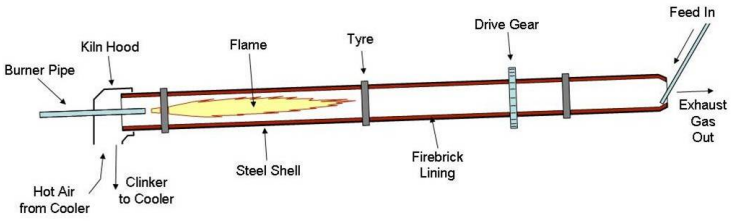


Figure 1.1: *Stages in a FEM Modeling Procedure.*

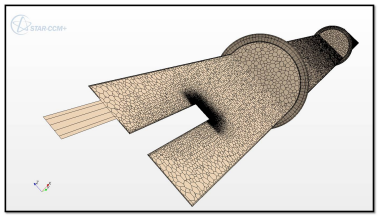
1.2 Motivating Examples

Industrial Furnace The simulation of flows through an enclosed surface is an often reoccuring problem in computational fluid dynamics. In Figure 1.2 a study case of the simulation of the flow and temperature profile inside an industrial furnace is presented.

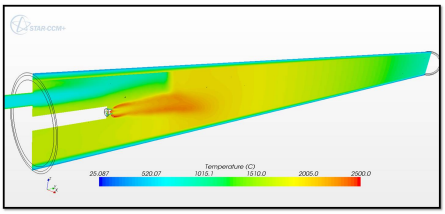
Fault current limiters Fault current limiters are expected to play an important role in the protection of future power grids. They are capable of preventing fault currents from reaching too high



(a) Schematic represention of an industrial furnace



(b) Mesh employed



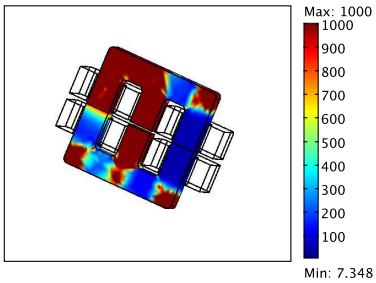
(c) Simulated temperature profile.

Figure 1.2: *Numerical simulation of industrial furnaces.*

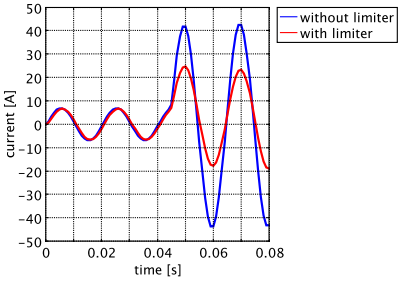
levels and, therefore enhance the life time expectancy all power system components. Figure 1.3 shows two examples of fault-current limiters along with some finite element simulation results.

(a) Open core configuration

(b) Three legged configuration



(c) Diffusion coefficient μ



(d) Current with and without limiter

Figure 1.3: *Numerical simulation of fault current limiters.*

1.3 Classification of Second Order Partial Differential Equations

In this section we classify second order linear partial differential equations (PDEs) with constant coefficients according to their elliptic, parabolic and hyperbolic nature and give examples of PDEs in each of these three classes.

Classification We consider an open two-dimensional domain $(x, y) \in \Omega \subset \mathbb{R}^2$ with boundary $\Gamma = \partial\Omega$ as the domain of the second order linear partial differential equation (PDE) for the unknown field $u = u(x, y)$ and the source function $f(x, y)$ that can be written as

$$\mathcal{L}(u) = f \text{ on } \Omega \tag{1.1}$$

where the operator $\mathcal{L}$ has constant coefficients a_{ij}, b_i and c

$$\mathcal{L}(u) = a_{11}\frac{\partial^2 u(x,y)}{\partial x^2} + 2a_{12}\frac{\partial^2 u(x,y)}{\partial x \partial y} + a_{22}\frac{\partial^2 u(x,y)}{\partial y^2} + b_1\frac{\partial u(x,y)}{\partial x} + b_2\frac{\partial u(x,y)}{\partial y} + c\,u(x,y)\,. \tag{1.2}$$

We classify these equation based on the sign of the determinant

$$D = \begin{vmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}^2\,. \tag{1.3}$$

The differential operator $\mathcal{L}$ is called

- *elliptic* if $D > 0$
- *parabolic* if $D = 0$
- *hyperbolic* if $D < 0$

Prototypes in this classification are the elliptic Laplace equation $u_{xx} + u_{yy} = 0$, parabolic heat equation $u_{xx} - u_y = 0$ and the hyperbolic wave equation $u_{xx} - u_{yy} = 0$. In the case of xy-varying coefficients a_{11}, a_{22} and a_{12}, the sign of D and therefore the type of the PDE may change with the location in Ω.

As FEM solvers for discretized hyperbolic PDEs are far less developed, we will only consider elliptic and parabolic equations in this course. Elliptic equations are characterized by the fact that changes in the data imposed on the boundary Γ are felt instantaneously in the interior of Ω. Parabolic equations model the evolution of a system from initial to an stationary (time-independent) equilibrium state.

Examples of Elliptic Partial Differential Equations We will consider the Poisson equation

$$-\triangle u := -\frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} = f \tag{1.4}$$

as it plays a central role in engineering applications in which the field u solved for plays the role of a (electric, magnetic or gravitational) potential, temperature or displacement. The reason for the minus sign in front of the Laplacian will be discussed together with the finite element

discretization. We will also consider two variants of (1.4). In the first variant we introduce a small positive parameter $0 < \epsilon \ll 1$ to arrive at the anisotropic variant

$$-\epsilon \frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} = f \tag{1.5}$$

that can be thought of as a simplified model to study the effect of local mesh refinement required to capture e.g. boundary layers or small geometrical details (such as for instance the air-gap in an electrical machine). In the second variant we again incorporate more physical realism in (1.4) by allowing the parameter c in

$$-\frac{\partial}{\partial x}\left(c\,\frac{\partial u(x,y)}{\partial x}\right) - \frac{\partial}{\partial y}\left(c\,\frac{\partial u(x,y)}{\partial y}\right) = f\,, \tag{1.6}$$

to have a jump-discontinuity across the boundary of two subdomains (such as for instance the jump in electrical conductivity across the interface between copper and air).

The Helmholtz and the convection-diffusion equation are other examples of elliptic partial differential equations. The former can be written as

$$-\triangle u - k^2 u = -\frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} - k^2 u = f\,, \tag{1.7}$$

and models the propagation of an acoustical or electromagnetic wave with wave number k. The latter is a simplified model for the study of a flow with (dimensionless) velocity $\mathbf{v} = (1,0)$ around a body and can be written as

$$-\epsilon \triangle u + \mathbf{v}\cdot\nabla u = -\epsilon \frac{\partial^2 u(x,y)}{\partial x^2} - \epsilon \frac{\partial^2 u(x,y)}{\partial y^2} + \frac{\partial u(x,y)}{\partial x} = f\,, \tag{1.8}$$

where ϵ is the inverse of the Peclet number (which plays the same role as the Reynolds number in the Navier-Stokes equations). The verification that the equations introduced above are indeed elliptic is left as an exercise.

The classification given above can be extended to systems of coupled partial differential equations. The vector-valued double curl equation for example relates the vector potential $\mathbf{U}(x,y,z) = (U_1(x,y,z), U_2(x,y,z), U_3(x,y,z))$ with the excitation vector field $\mathbf{F}(x,y,z)$ and can be written as

$$\nabla \times (\nu \nabla \times \mathbf{U}) = \mathbf{F}(x,y,z)\,, \tag{1.9}$$

where as in (1.6) ν represents a material parameter. This model can be derived from the Maxwell equations and plays a central role in the finite element modeling of electrical machines, wind turbines and power transformers. In this setting $\mathbf{F}$ and $\mathbf{U}$ play the role of the applied current density and the magnetic vector potential, respectively.

In order to guarantee uniqueness of the solution of the above elliptic partial differential equations, appropriate (Dirichlet, Neumann or other) boundary conditions on $\Gamma = \partial\Omega$ need to be supplied.

Example of a Parabolic Partial Differential Equation In order to give examples of parabolic partial differential equations, we change the notation y into t and consider Ω to be a rectangle $(x,t) \in \Omega = [0,L] \times [0,T]$. The parabolic differential equation with source term $f(x,t)$

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial^2 u(x,t)}{\partial x^2} + f(x,t) \tag{1.10}$$

models the diffusion of the quantity $u(x,t)$ in time that reaches a time-independent equilibrium state when $\frac{\partial u(x,t)}{\partial t}$ is zero. All of the elliptic models introduced above can be extended to a parabolic model by adding the allowing u to become time-dependent, that is $u = u(x,y,t)$, and adding the term $\frac{\partial u(x,t)}{\partial t}$. These models will be considered for instance when discussing the role of initial vectors of iterative solution methods.

In order to guarantee uniqueness of the solution of parabolic partial differential equations, both boundary and initial conditions need to be supplied.

1.4 Model Problems

In this section we define a sequence of model problems for future reference.

Elliptic Problem Problems

- given the domain $x \in \Omega = (a,b) \subset \mathbb{R}$, given a positive diffusion coefficient $c(x) > 0$, $\forall x \in (a,b)$, and given the boundary data $g(x)$ or $h(x)$, solve for $u(x)$ the following ordinary differential equation

$$-\frac{d^2u(x)}{dx^2} = f(x) \text{ on } \Omega \tag{1.11}$$

or

$$-\frac{d}{dx}\left(c(x)\,\frac{du(x)}{dx}\right) = f(x) \text{ on } \Omega \tag{1.12}$$

supplied with either Dirichlet boundary conditions

$$u(x) = g(x) \text{ on } x = a \text{ and/or } x = b \tag{1.13}$$

or Neumann boundary conditions

$$\frac{d\,u(x)}{dx} = h(x) \text{ on } \text{ on } x = a \text{ and/or } x = b\,. \tag{1.14}$$

In case that $g(x) = 0$ in (1.13) ($h(x) = 0$ in (1.14)) the Dirichlet (Neumann) boundary conditions is called homogeneous.

- given an open and bounded domain $(x,y) \in \Omega \subset \mathbb{R}^2$ with bounded by a close curve $\Gamma = \partial\Omega$ that can be subdivided as $\Gamma = \Gamma_D \cup \Gamma_N$ and has outward normal $\mathbf{n}$, given a positive diffusion coefficient $c(x,y) > 0$, $\forall (x,y) \in \Omega$, and given the boundary data $g(x,y)$ or $h(x,y)$, solve for $u(x,y)$ the following partial differential equation

$$-\triangle u := -\frac{\partial^2 u(x,y)}{\partial x^2} - \frac{\partial^2 u(x,y)}{\partial y^2} = f(x,y) \text{ on } \Omega\,, \tag{1.15}$$

or

$$-\frac{\partial}{\partial x}\left(c(x,y)\,\frac{\partial u(x,y)}{\partial x}\right) - \frac{\partial}{\partial y}\left(c(x,y)\,\frac{\partial u(x,y)}{\partial y}\right) = f(x,y) \text{ on } \Omega\,, \tag{1.16}$$

supplied with either Dirichlet boundary conditions

$$u(x,y) = g(x,y) \text{ on } \Gamma_D \tag{1.17}$$

or Neumann boundary conditions

$$\frac{\partial u(x,y)}{\partial n} = \nabla u(x,y) \cdot \mathbf{n} = h(x,y) \text{ on } \Gamma_N\,. \tag{1.18}$$

- given an open and bounded domain $(x, y, z) \in \Omega \subset \mathbb{R}^3$ bounded by a closed surface $S = \partial\Omega$ that can be subdivided into $S = S_S \cup S_N$ and has outward normal $\mathbf{n}$, given a positive coefficient $\nu(x, y, z) > 0$, $\forall (x, y, z) \in \Omega$, given a source function $\mathbf{F}(x, y, z)$ and given the boundary data $\mathbf{G}(x, y, z)$ or $\mathbf{H}(x, y, z)$, solve for $U(x, y, z)$ the following partial differential

$$\nabla \times (\nu \nabla \times \mathbf{U}) = \mathbf{F}(x, y, z)\,, \tag{1.19}$$

supplied with either Dirichlet boundary conditions

$$\mathbf{U}(x, y, z) = \mathbf{G}(x, y, z) \text{ on } S_D\,, \tag{1.20}$$

or conditions on the *tangential* derivatives of $\mathbf{U}$

$$(\nabla \times \mathbf{U}) \times \mathbf{n} = \mathbf{G}(x, y, z) \text{ on } S_N\,. \tag{1.21}$$

Parabolic Model Problems

- given the domain $x \in \Omega = (a, b) \subset \mathbb{R}$ and the time time interval $t \in [0, T]$, given a positive diffusion coefficient $c(x) > 0$, $\forall x \in (a, b)$, given initial conditions $u_0(x) = u(x, t = 0)$ and given the boundary data $g(x)$ or $h(x)$, solve for $u(x, t)$ the following partial differential equation

$$\frac{\partial u(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(c(x)\,\frac{\partial u(x,t)}{\partial x}\right) + f(x) \text{ on } \Omega \times [0, T] \tag{1.22}$$

supplied with either Dirichlet boundary conditions

$$u(x, t) = g(x) \text{ on } x = a \text{ and/or } x = b \quad \forall t \in [0, T] \tag{1.23}$$

or Neumann boundary conditions

$$\frac{\partial\, u(x,t)}{\partial x} = h(x) \text{ on } \text{ on } x = a \text{ and/or } x = b \quad \forall t \in [0, T] \tag{1.24}$$

and the initial condition

$$u(x, t = 0) = u_0(x) \quad \forall x \in \Omega\,. \tag{1.25}$$

2

Mathematical Preliminaries

The goals and motivations of this chapter are to

- generalize the concept of a primitive of a scalar function in a single real variable to vector-valued functions resulting in the concept of a (scalar or vector) potential;
- generalize the integration by parts formula for integrals of scalar functions to integrals of vector-valued functions allowing to cast the strong form of a partial differential equation problem into its weak (or variational) form;
- give the fundamental lemma of variational calculus as allowing to demonstrate the equivalence of the strong and weak form of a partial differential equation problem.

We refer to Stuart for details.

2.1 The Geometry of $\mathbb{R}^3$

We denote a point in the three-dimensional Euclidean space $\mathbb{R}^3$ by $\mathbf{x} = (x, y, z)$. We denote a vector $\mathbf{u}$ in $\mathbb{R}^3$ either by its components (u_1, u_2, u_3) or by the (vectorial) sum of the components times the respective unit vectors in x, y and z-direction, .i.e, $\mathbf{u} = u_1\,\mathbf{i} + u_2\,\mathbf{j} + u_3\,\mathbf{k}$.

Inner product and norm We denote the *inner product* (or scalar product) of two vectors $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ in $\mathbb{R}^3$ by $\mathbf{u} \cdot \mathbf{v}$. The result of this product is a number (or scalar) defined as

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + u_3 v_3\,. \tag{2.1}$$

The *norm* of the vector $\mathbf{u}$ denoted as $\|\mathbf{u}\|$ is defined as

$$\|\mathbf{u}\| = \sqrt{\mathbf{u} \cdot \mathbf{u}} = \sqrt{u_1 u_1 + u_2 u_2 + u_3 u_3}\,. \tag{2.2}$$

Outer product We denote the *outer product* (or vector product) of two vectors $\mathbf{u} = (u_1, u_2, u_3)$ and $\mathbf{v} = (v_1, v_2, v_3)$ in $\mathbb{R}^3$ by $\mathbf{u} \times \mathbf{v}$. The result of this product is a vector defined as

$$\mathbf{u} \times \mathbf{v} = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} = [u_2 v_3 - u_3 v_2]\ \mathbf{i} - [u_1 v_3 - u_3 v_1]\,\mathbf{j} + [u_1 v_2 - u_2 v_1]\,\mathbf{k}\,. \tag{2.3}$$

The *triple product* of three vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ is defined as $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w}$ and satisfies

$$(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{w} = \mathbf{u} \cdot (\mathbf{v} \times \mathbf{w}) \tag{2.4}$$

2.2 Calculus of Functions of a Real Variable

Assume $f(x)$ to be a scalar function in a single real variable x. A point x_0 in the domain of f such that $\frac{df}{dx}(x = x_0) = f'(x_0) = 0$ is called a critical point of $f(x)$. The function $f(x)$ attains its (local) minima and maxima in its first order critical points (though not all first order critical points are minima or maxima of $f(x)$). The Fundamental theorem of Calculus states that, given a function $f(x)$ in a single real variables with primitive $F(x)$, the integral $\int_a^b f(x)\,dx$ can be computed as

$$\int_a^b f(x)\,dx = F(b) - F(a)\,. \tag{2.5}$$

The following integration by parts formula holds

$$\int_a^b f(x)\,g'(x)\,dx = [f(x)g(x)]_a^b - \int_a^b f'(x)\,g(x)\,dx\,. \tag{2.6}$$

2.3 Calculus of Functions of Several Variables

Assume $u(\mathbf{x}) = u(x, y, z)$ to be be a function on $\mathbb{R}^3$, i.e., $u(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}$. Assume furthermore $\mathbf{x}_0$ to be point in the domain of $u(\mathbf{x})$ and $\mathbf{s}$ a unit vector, respectively.

Directional Derivative The derivative of $u(\mathbf{x})$ in the direction $\mathbf{s}$ in the point $\mathbf{x}_0$ is denoted as $D_{\mathbf{s}}u(\mathbf{x}_0)$ and defined as

$$D_{\mathbf{s}}u(\mathbf{x}_0) = \lim_{\epsilon \to 0} \frac{u(\mathbf{x}_0 + \epsilon\,\mathbf{s}) - u(\mathbf{x}_0)}{\epsilon} = \frac{d}{d\epsilon}\left[u(\mathbf{x}_0 + \epsilon\,\mathbf{s})\right]|_{\epsilon=0}\,. \tag{2.7}$$

This derivative gives the rate of change $u(\mathbf{x}) = u(x, y, z)$ in the direction of $\mathbf{s}$ at the point $\mathbf{x}_0$.

Gradient The *gradient* of u denoted as $\operatorname{grad} u = \nabla u$ is vector function that is defined as

$$\nabla u = \left(\frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, \frac{\partial u}{\partial z}\right) = \frac{\partial u}{\partial x}\,\mathbf{i} + \frac{\partial u}{\partial y}\,\mathbf{j} + \frac{\partial u}{\partial z}\,\mathbf{k} = u_x\,\mathbf{i} + u_y\,\mathbf{j} + u_z\,\mathbf{k}\,. \tag{2.8}$$

The gradient of u is allows to compute the directional derivative of u in the direction of $\mathbf{s}$. We indeed have that $D_{\mathbf{s}}u(\mathbf{x}_0) = \nabla u(\mathbf{x}_0) \cdot \mathbf{s}$. This identity will be useful in formulating necessary conditions for first order critical points of $u(\mathbf{x})$. In a physical context and in case that u is e.g. the temperature, ∇u is proportional to the heat flux.

First Order Critical Points The point $\mathbf{x}_0$ is a first order critical point of $u(\mathbf{x}_0)$ iff $\nabla u(\mathbf{x}_0) = \mathbf{0}$ in $\mathbb{R}^3$. The latter condition can be stated equivalently as

$$\frac{d}{d\epsilon}\left[u(\mathbf{x}_0 + \epsilon\,\mathbf{s})\right]|_{\epsilon=0} = 0\,(\text{ in } \mathbb{R}) \quad \forall \mathbf{s} \in \mathbb{R}^3 \text{ with } \|\mathbf{s}\| = 1\,. \tag{2.9}$$

This condition will be allow to define first order critical point of functionals on infinite dimensional vector spaces in the next chapter.

2.4 Calculus of Vector Functions

Assume $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}), F_3(\mathbf{x}))$ to be a vector function on $\mathbb{R}^3$, i.e., $\mathbf{F}(\mathbf{x}) : \mathbb{R}^3 \mapsto \mathbb{R}^3$.

Divergence and curl The *divergence* of $\mathbf{F}$ denoted as $\operatorname{div}\mathbf{F} = \nabla \cdot \mathbf{F}$ is a scalar function that is defined as

$$\nabla \cdot \mathbf{F} = \frac{\partial F_1}{\partial x} + \frac{\partial F_2}{\partial y} + \frac{\partial F_3}{\partial z}\,. \tag{2.10}$$

The *curl* (or rotation) of $\mathbf{F}$ denoted as $\operatorname{curl}\mathbf{F} = \nabla \times \mathbf{F}$ is a vector function that is defined as

$$\nabla \times \mathbf{F} = \begin{pmatrix} \frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z} \\ -\frac{\partial F_3}{\partial x} + \frac{\partial F_1}{\partial z} \\ \frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y} \end{pmatrix} = \left(\frac{\partial F_3}{\partial y} - \frac{\partial F_2}{\partial z}\right)\mathbf{i} - \left(\frac{\partial F_3}{\partial x} - \frac{\partial F_1}{\partial z}\right)\mathbf{j} + \left(\frac{\partial F_2}{\partial x} - \frac{\partial F_1}{\partial y}\right)\mathbf{k}\,. \tag{2.11}$$

The divergence of the vector field F evaluated at a particular point $\mathbf{x}$ is the extent to which the vector field behaves like a source or a sink at a given point. If the divergence is nonzero at some point then there must be a source or sink at that position. The curl of a vector field $\mathbf{F}$ describes an infinitesimal rotation of $\mathbf{F}$.

Example 2.4.1. *Insert examples as in previous version of the ATHENS notes here.*

Example 2.4.2. *Assume $c(\mathbf{x})$ and $u(\mathbf{x})$ to be two scalar functions. Then $c\,\nabla u$ is a vector function whose divergence is equal to*

$$\nabla \cdot (c\,\nabla u) = \frac{\partial}{\partial x}\left(c\,\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(c\,\frac{\partial u}{\partial y}\right) + \frac{\partial}{\partial z}\left(c\,\frac{\partial u}{\partial z}\right)\,. \tag{2.12}$$

If in particular $c = 1$ (i.e. $c(\mathbf{x}) = 1$ for all $\mathbf{x}$), then

$$\nabla \cdot \nabla u = \triangle\phi = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2}\,, \tag{2.13}$$

where $\triangle u$ denotes the Laplacian *of u. This differential operator is generalized to vector-valued functions $\mathbf{F}(\mathbf{x})$ by the* double-curl *operator as $\nabla \times (c\nabla \times \mathbf{F})$.*

Proposition 2.4.1. *From the above definitions immediately follows that for any scalar function u and for any vector function $\mathbf{F}$ that*

$$\nabla \times (\nabla u) = \mathbf{0}\ (in\ \mathbb{R}^3) \tag{2.14}$$

$$\nabla \cdot (\nabla \times \mathbf{F}) = 0\ (in\ \mathbb{R})\,. \tag{2.15}$$

We will see below that scalar functions u that satisfy $\nabla \cdot u = 0$ and vector functions $\mathbf{F}$ that satisfy $\nabla \times \mathbf{F} = \mathbf{0}$ play a special role as they generalize the notion of a constant function $f(x)$ for which $f'(x) = 0$.

Proposition 2.4.2. *Let c, u and v be scalar functions and let $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{U}$ be vector functions. By applying the product rule for differentiation, the following properties can be derived*

$$\nabla \cdot (v\mathbf{F}) = v\nabla \cdot \mathbf{F} + \nabla v \cdot \mathbf{F} \tag{2.16}$$

$$\nabla \cdot (\mathbf{F} \times \mathbf{G}) = (\nabla \times \mathbf{F}) \cdot \mathbf{G} + (\nabla \times \mathbf{G}) \cdot \mathbf{F}\,. \tag{2.17}$$

Setting in the former equality $\mathbf{F} = c\,(\nabla u)$, yields

$$\nabla \cdot [(c\,\nabla u)v] = [\nabla \cdot (c\,\nabla u)]\,v + c\,\nabla u \cdot \nabla v\,. \tag{2.18}$$

while putting in the latter equality $\mathbf{F} = c\,(\nabla \times \mathbf{U})$, yields

$$\nabla \cdot [c\,(\nabla \times \mathbf{U}) \times \mathbf{G}] = [\nabla \times (c\nabla \times \mathbf{U})] \cdot \mathbf{G} + (\nabla \times \mathbf{G}) \cdot (c\nabla \times \mathbf{U})\ . \tag{2.19}$$

Flux and first integration theorem Assume S to be an oriented surface in $\mathbb{R}^3$ with unit outward normal $\mathbf{n}$. We will denote by $-S$ the surface that coincides with S and has opposite orientation. The normal derivative on S of the function u denoted as u_n is defined as $u_n = \nabla u \cdot \mathbf{n}$. The *flux* of $\mathbf{F}$ through S denoted as Φ is defined as the surface integral

$$\Phi = \int_S \mathbf{F} \cdot d\mathbf{S} = \int_S \mathbf{F} \cdot \mathbf{n}\, dS\,. \tag{2.20}$$

Assume the surface S to be closed and enclosing a volume V. We make this relation explicit by writing $S = S(V)$. The Gauss integration theorem states that

$$\int_{S(V)} \mathbf{F} \cdot d\mathbf{S} = \int_V \nabla \cdot \mathbf{F}\, dV\,. \tag{2.21}$$

Proposition 2.4.3. *Assume as in Proposition 2.4.2 c, u and v to be scalar functions and $\mathbf{F}$, $\mathbf{G}$ and $\mathbf{U}$ to be vector functions. Applying Gauss' Theorem to the vector function $v\,\mathbf{F}$ and using (2.16) to expand the right-hand side yields*

$$\int_{S(V)} \mathbf{F}\, v \cdot d\mathbf{S} = \int_V \nabla \cdot [\mathbf{F}\, v]\ dV = \int_V v\nabla \cdot \mathbf{F}\, dV + \int_V \nabla v \cdot \mathbf{F}\, dV\,. \tag{2.22}$$

Setting $\mathbf{F} = c\,(\nabla u)$ yields

$$\int_{S(V)} [c\,(\nabla u)]\ v \cdot d\mathbf{S} = \int_V v\nabla \cdot [c\,(\nabla u)]\ dV + \int_V \nabla v \cdot [c\,(\nabla u)]\ dV\,. \tag{2.23}$$

Using the notation $u_n = \nabla u \cdot \mathbf{n}$ for the normal derivative, one can write the integrand in the left-hand side as $[c\,(\nabla u)]\ v \cdot d\mathbf{S} = [c\,(\nabla u)]\ v \cdot \mathbf{n}\, dS = c\, u_n\, v\, dS$. After rearranging terms, we obtain the following integration by parts formula

$$\int_V [\nabla \cdot (c\,\nabla u)]\, v\, dV = \int_{S(V)} c\, u_n\, v\, dS - \int_V c\,\nabla u \cdot \nabla v\, dV\,. \tag{2.24}$$

Applying Gauss' Theorem to the vector function $\mathbf{F} \times \mathbf{G}$ and using (2.17) to expand the right-hand side yields

$$\int_{S(V)} \mathbf{F} \times \mathbf{G} \cdot d\mathbf{S} = \int_V \nabla \cdot (\mathbf{F} \times \mathbf{G})\ dV = \int_V (\nabla \times \mathbf{F}) \cdot \mathbf{G}\, dV + \int_V (\nabla \times \mathbf{G}) \cdot \mathbf{F}\, dV\,. \tag{2.25}$$

We use the identify (2.4) to rewrite the integrand in the left-hand side as $\mathbf{F}\times\mathbf{G}\cdot d\mathbf{S} = (\mathbf{F}\times\mathbf{G})\cdot\mathbf{n}\,dS = \mathbf{F}\cdot(\mathbf{G}\times\mathbf{n})\,dS$. *Setting* $\mathbf{F} = c\,(\nabla\times\mathbf{U})$ *and rearranging terms then yields*

$$\int_V [\nabla\times(c\,\nabla\times\mathbf{U})]\cdot\mathbf{G}\,dV = \int_{S(V)} c\,(\nabla\times\mathbf{U})\cdot(\mathbf{G}\times\mathbf{n})\,dS - \int_V (c\,\nabla\times\mathbf{U})\cdot(\nabla\times\mathbf{G})\;dV\,. \quad (2.26)$$

In subsequent chapters, we will extensively use a two-dimensional variant of (2.24) in which an oriented surface Ω with outward normal $\mathbf{n}$ in $\mathbb{R}^2$ (that later we will later call a computational domain) is bounded by a boundary $\Gamma = \partial\Omega$. In this case we the following integration by parts formula

$$\boxed{\int_\Omega [\nabla\cdot(c\,\nabla u)]\,v\,d\Omega = \int_{\Gamma(\Omega)} c\,u_n\,v\,d\Gamma - \int_\Omega c\,\nabla u\cdot\nabla v\,d\Omega\,,} \quad (2.27)$$

Divergence-Free Vector Fields and Existence of Vector Potentials A vector function $\mathbf{F}$ is called *divergence-free* (or solenoidal) on $\Omega\subset\mathbb{R}^3$ if and only if $\nabla\cdot\mathbf{F}(\mathbf{x}) = 0\,(\text{in }\mathbb{R})$ for all $\mathbf{x}\in\Omega$. Unless stated differently, we will call in these notes a vector function $\mathbf{F}$ divergence-free if and only if it is divergence-free on $\Omega = \mathbb{R}^3$.

Gauss' theorem implies that the flux of a divergence-free vector field through a closed surface S is zero. Decomposing S as $S = S^+ + (-S^-)$, the flux of a divergence-free $\mathbf{F}$ through S^+ and S^- can then be seen to be equal. Given that S was chosen arbitrarily, the flux of a divergence-free vector in *surface independent*. This surface-independent property allows in turn to demonstrate that for divergence-free vector fields a vector potential $\mathbf{A}$ must exist such that

$$\boxed{\mathbf{F} = \nabla\times\mathbf{A}}\,, \quad (2.28)$$

or component-wise

$$F_1 = \left(\frac{\partial A_3}{\partial y} - \frac{\partial A_2}{\partial z}\right), F_2 = \left(-\frac{\partial A_3}{\partial x} + \frac{\partial A_1}{\partial z}\right) \text{ and } F_3 = \left(\frac{\partial A_2}{\partial x} - \frac{\partial A_1}{\partial y}\right). \quad (2.29)$$

Circulation and second integration theorem Assume Γ to be a *closed* oriented curve in $\mathbb{R}^3$ with unit tangential vector $\mathbf{t}$. We will denote by $-\Gamma$ the curve that coincides with Γ and has opposite orientation. The tangential derivative along Γ of the function u denoted as u_t is defined as $u_t = \nabla u\cdot\mathbf{t}$. The *circulation* of $\mathbf{F}$ along Γ denoted as $\mathcal{C}$ is defined as the line integral

$$\mathcal{C} = \int_\Gamma \mathbf{F}\cdot d\mathbf{\Gamma} = \int_\Gamma \mathbf{F}\cdot\mathbf{t}\,d\Gamma\,. \quad (2.30)$$

Assume that the curve Γ encloses a surface S and that this relation in written as $\Gamma(S)$. The Stokes' integration theorem states that

$$\int_{\Gamma(S)} \mathbf{F}\cdot d\mathbf{\Gamma} = \int_S (\nabla\times\mathbf{F})\cdot d\mathbf{S}\,. \quad (2.31)$$

Curl-Free Vector Fields and Existence of Scalar Potentials A vector function $\mathbf{F}$ is called *curl-free* (or irrotational) on $\Omega\subset\mathbb{R}^3$ if and only if $\nabla\times\mathbf{F}(\mathbf{x}) = \mathbf{0}(\text{in }\mathbb{R}^3)$ for all $\mathbf{x}\in\Omega$. Unless stated differently, we will call in these notes a vector function $\mathbf{F}$ curl-free if and only if it is curl-free on $\Omega = \mathbb{R}^3$.

Stokes' theorem implies that the circulation of a curl-free vector field along a closed curve Γ is zero. Decomposing Γ as $\Gamma = \Gamma^+ + (-\Gamma^-)$, the line integral of a divergence-free $\mathbf{F}$ along Γ^+ and Γ^- can then be seen to be equal. Given that Γ was chosen arbitrarily, the circulation of a curl-free vector is *path independent*. This path-independent property allows in turn to demonstrate that for curl-free vector fields a scalar potential u must exist such that

$$\boxed{\mathbf{F} = \nabla \cdot u}\,, \tag{2.32}$$

or component-wise

$$F_1 = \frac{\partial u}{\partial x}\,, F_2 = \frac{\partial u}{\partial y} \text{ and } F_3 = \frac{\partial u}{\partial z}\,. \tag{2.33}$$

3

Minimization Problems and Differential Equations

A physical system evolves in such a way to minimize their internal energy. Newton's apple falls due to gravitation. The thermal or magnetic flux passes by preference through regions of high (thermal or magnetic) conductivity. Partial differential equations are derived from this principle. Instead of solving the differential problem directly it might therefore be more advantegeous to consider an equivalent energy minimization problem. In this chapter we put the equivalence between partial differential equations and minimization problems in a more formal setting as it will serve as a tool to explain FEM to solve the former. In this chapter we aim at

[1] give the fundamental lemma of variational calculus known as du Bois-Reymond's lemma;

[2] derive the Euler-Lagrange equation whose solution describes a stationary point for a functional to be minimized;

[3] use the above two ingredients to show how partial differential equations can be derived from a minimization problem;

[4] show under which conditions a partial differential differential can be derived from a minimization problem.

3.1 Function Spaces, Differential Operators and Functionals

In this section we introduce ingredients allowing to describe minimization problems formally. We seek to describe that the orbit that the falling apple describes is an orbit that minimizes its energy.

■ 3.1.1 Function Spaces

In the following we will denote Ω as open and bounded domain in $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$ that we will refer to as the computational domain. As Ω is open, it does not contain its boundary. The union of Ω

and its boundary is called the closure of Ω and denoted as $\overline{\Omega}$.

We recall that a space V is a *vector space* iff V is closed for linear combinations, iff $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2 \in V$ implies that $c_1 v_1 + c_2 v_2 \in V$.

Definition 3.1.1. (Function space) *Given an open and bounded domain Ω (in $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$), a function space $V(\Omega)$ is a vector space of functions (in one, two or three variables) on Ω that satisfy some smoothness requirements and possibly boundary conditions.*

The set $V(\Omega)$ is e.g. the set of all imaginable orbits of the Newton's apple. We impose sufficient smoothness conditions to allow to take derivatives of elements in $V(\Omega)$. If additionally boundary conditions are imposed, elements $V(\Omega)$ can be viewed as solutions of a partial differential equation supplied with boundary conditions. More formally, we can give the following examples:

Example 3.1.1. (One-Dimensional Domain) *Given the open interval $\Omega = (a, b) \subset \mathbb{R}$, we will denote by $L^2(\Omega)$ the set of square integrable functions on Ω, i.e., $u \in L^2(\Omega)$ iff $\int_a^b u^2(x)\,dx < \infty$ and by $H^1(\Omega)$ the set of square integrable function with square integrable first derivative, i.e, $u \in H^1(\Omega)$ iff $\int_a^b u^2(x)\,dx < \infty$ and $\int_a^b \left(\frac{du}{dx}\right)^2 dx < \infty$. Clearly $H^1(\Omega) \subset L^2(\Omega)$. We will denote by $\mathcal{C}(\Omega)$ the function that are continuous on $\overline{\Omega}$. If $u \in \mathcal{C}(\Omega)$, than u is bounded (as continuous function function on a compact set) and therefore square integrable, i.e., $\mathcal{C}(\Omega) \subset L^2(\Omega)$. We will denote by $\mathcal{C}^1(\Omega)$ and $\mathcal{C}^2(\Omega)$ the functions $u(x)$ with a continuous first derivative $u'(x)$ and continuous second derivative $u''(x)$ on $\overline{\Omega}$, respectively.*

Assume that $x_0 \in (a, b)$ and $\delta > 0$ such that both $x_{-\delta} = x_0 - \delta \in (a, b)$ and $x_\delta = x_0 + \delta \in (a, b)$. We define the functions $v_1(x)$, $v_2(x)$ and $v_3(x)$ as piecewise constant, linear and quadratic polynomials follows

$$v_1(x) = \begin{cases} 0 \text{ if } |x - x_{-\delta}| > \delta \\ 1 \text{ if } |x - x_{-\delta}| \leq \delta \end{cases} \tag{3.1}$$

$$v_2(x) = \begin{cases} 0 \text{ if } |x - x_{-\delta}| > \delta \\ (x - x_{-\delta})/(x_0 - x_{-\delta}) \text{ if } x_{-\delta} \leq x \leq x_0 \\ (x_\delta - x)/(x_\delta - x_0) \text{ if } x_0 \leq x \leq x_\delta \end{cases} \tag{3.2}$$

$$v_3(x) = \begin{cases} 0 \text{ if } |x - x_{-\delta}| > \delta \\ (x - x_{-\delta})^2/(x_0 - x_{-\delta})^2 \text{ if } x_{-\delta} \leq x \leq x_0 \\ (x_\delta - x)^2/(x_\delta - x_0)^2 \text{ if } x_0 \leq x \leq x_\delta \end{cases} \tag{3.3}$$

Then $v_1(x)$ has a jump discontinuity in both $x = x_{-\delta}$ and $x = x_\delta$, and therefore $v_1(x) \notin \mathcal{C}(\Omega)$. The function $v_2(x)$ is continuous, but its derivative has a jump discontinuity in $x = x_{-\delta}$, $x = x_0$ and $x = x_\delta$, and therefore $v_2(x) \in \mathcal{C}(\Omega) \setminus \mathcal{C}^1(\Omega)$. The function $v_3(x)$ is continuous and has a continuous first derivative, but its second derivative has a jump discontinuity in $x = x_{-\delta}$, $x = x_0$ and $x = x_\delta$, and therefore $v_3(x) \in \mathcal{C}^1(\Omega) \setminus \mathcal{C}^2(\Omega)$. The function $v_2(x)$ is informly referred to as the hat-function, and will play a central role in the application of the FEM method.

We denote the subset $\mathcal{C}_0(\Omega) = \{u \in \mathcal{C}(\Omega) \mid u(a) = 0 = u(b)\}$ of functions satisfying homogeneous Dirichlet boundary conditions. The space $\mathcal{C}_0^1(\Omega)$ and $\mathcal{C}_0^2(\Omega)$ are defined similarly.

Example 3.1.2. (Two-Dimensional Domain) *In problems on two-dimensional domain, we will denote by $\Omega \subset \mathbb{R}^2$ an open and bounded domain with boundary $\Gamma = \partial\Omega$ that can be partioned into distinct parts Γ_D and Γ_N, i.e., $\Gamma = \Gamma_D \cup \Gamma_N$. We associate with Γ_D and Γ_N that part of the boundary on which (possibly non-homogeneous) Dirichlet and Neumann (or Robin) boundary conditions are*

imposed, respectively. We will denote by $\mathcal{C}^1(\Omega)$ *the functions* $u(x,y)$ *with a continuous first order partial derivatives* u_x *and* u_y*. We will denote the subset* $\mathcal{C}_0^1(\Omega) = \left\{u \in \mathcal{C}^1(\Omega) \mid u_{|\Gamma_D} = 0\right\}$.

Treatment of boundary conditions The presence of the Neumann boundary conditions is not reflected in the above definition of $\mathcal{C}_0^1(\Omega)$. The condition $u_{|\Gamma_D} = 0$ is imposed even in case of non-homogenous Dirichlet boundary conditions. Only with the latter condition the above function spaces are vector space (closed for linear combinations) of infinite dimensions. These vector spaces can be endowed with a inner-product, thus becoming therefore Hilbert spaces.

■ 3.1.2 Differential Operators

Assume that $\mathcal{L}$ is a second order differential operator acting on elements of $V(\Omega)$ and let $u = 0$, i.e., $u(\mathbf{x}) = 0\ \forall \mathbf{x} \in \Omega$, denote the zero element of $V(\Omega)$. It is obvious that $\mathcal{L}u = 0$. This trivial situation is excluded in the following definition.

Definition 3.1.2. (Positivity) *The differential operatore* $\mathcal{L}$ *is positive on* $V(\Omega)$ *iff*

$$\int_\Omega v\,(\mathcal{L}v)\,d\Omega > 0 \qquad \forall v \in V(\Omega) \setminus \{0\}\ . \tag{3.4}$$

Observe that in the above definition the boundary conditions are taken into account by incorporating in the definition of $V(\Omega)$ and requiring that $v \in V(\Omega)$.

Example 3.1.3. *The differential operator* $\mathcal{L}v = -\frac{d^2v}{dx^2} = -v''(x)$ *is positive on* $\mathcal{C}_0^1((a,b))$*. Indeed, using integration by parts and the homogeneous Dirichtlet boundary conditions we have that*

$$\int_\Omega v\,(\mathcal{L}v)\,d\Omega = -\int_a^b v(x)\,v''(x)\,dx = -\underbrace{\left[v(x)\,v'(x)\right]_a^b}_{=0} + \int_a^b v'(x)\,v'(x)\,dx > 0\,. \tag{3.5}$$

Example 3.1.4. *Given on open and bounded subset* $\Omega \subset \mathbb{R}^2$ *and the coefficient* $c(\mathbf{x}) > 0\ \forall \mathbf{x} \in \Omega$*, the differential operator* $\mathcal{L}v = -\nabla \cdot (c\nabla v)$ *is positive on* $\mathcal{C}_0^1(\Omega)$*.*

Definition 3.1.3. (Self-adjointness) *The differential operator* $\mathcal{L}$ *is called self-adjoint (or symmetric) on* $V(\Omega)$ *iff*

$$\int_\Omega v\,(\mathcal{L}u)\,d\Omega = \int_\Omega u\,(\mathcal{L}v)\,d\Omega \qquad \forall u, v \in V(\Omega)\,. \tag{3.6}$$

Example 3.1.5. *The differential operator* $\mathcal{L}v = -\frac{d^2v}{dx^2}$ *is self-adjoint on* $\mathcal{C}_0^1((a,b))$*. This can be seen by integrating by parts twice and using the homogeneous boundary conditions.*

Example 3.1.6. *Given an open and bounded subset* $\Omega \subset \mathbb{R}^2$ *and the coefficient* $c(\mathbf{x}) > 0\ \forall \mathbf{x} \in \Omega$*, the differential operator* $\mathcal{L}v = -\nabla \cdot (c\nabla v)$ *is self-adjoint on* $\mathcal{C}_0^1(\Omega)$*.*

■ 3.1.3 Functionals

Definition 3.1.4. (Functional) *Given a function space* $V(\Omega)$*, a functional* L *is a mapping from* $V(\Omega)$ *onto* $\mathbb{R}$*, i.e.,* $L : V(\Omega) \mapsto \mathbb{R}$*.*

If $u \in V(\Omega)$ is an orbits of Newton's apple, then $L(u)$ can for instance represent the length of the orbit. If $u \in V(\Omega)$ is a temperature distribution in a region, then can for instance represent the average of the temperature. In the modeling of electrical machines the function u typically represents the (scalar or vector) magnetic potential and the force, torque, resistive heat and eddy current are examples of functionals in u. More formal examples are given below.

Example 3.1.7. (Distance in Flat Land) *Let $V(\Omega)$ be the set of differentiable functions on the open interval (a, b) supplied with the additional conditions that $u(x = a) = u_a$ and $u(x = b) = u_b$. The graph of each element of $V(\Omega)$ is a path between the points (a, u_a) and (b, u_b) in $\mathbb{R}^2$. The length $L(u)$ of each of these path can be computed by the line integral*

$$L(u) = \int ds \tag{3.7}$$

where, given that $dy = u'(x)\,dx$,

$$ds^2 = dx^2 + dy^2 = \left[1 + u'(x)^2\right] dx^2 \Leftrightarrow ds = \sqrt{1 + u'(x)^2}\,dx\,, \tag{3.8}$$

and therefore

$$L(u) = \int_a^b \sqrt{1 + u'(x)^2}\,dx\,. \tag{3.9}$$

Example 3.1.8. (Mass-Spring System) *Let $V(\Omega)$ be the set of differentiable functions on the open interval (a, b) describing the motion of a point (apple) with mass m vertically attached to a wall with a spring constant k (the independent variable $x \in (a, b)$ thus denotes time here). The instantaneous energy $E[u(x)]$ of the mass is the sum of the kinetic and potential energy $T[u(x)] = {}^1\!/_2\, m\, u'(x)^2$ and $V[u(x)] = {}^1\!/_2\, k\, u(x)^2$, i.e, $E[u(x)] = T[u(x)] - V[u(x)]$. The total energy is the functional $L(u)$ defined on $V(\Omega)$ as*

$$L(u) = \int_a^b \left[T[u(x)] - V[u(x)]\right] dx = \int_a^b \left[{}^1\!/_2\, m\, u'(x)^2 - {}^1\!/_2\, k\, u(x)^2\right] dx\,. \tag{3.10}$$

Example 3.1.9. (Abstract One-Dimensional Example) *The previous example can be put in a more abstract setting by allowing $f(x, u, u')$ to be a function in three variables mapping from $(a, b) \times V(\Omega) \times V(\Omega)$ to $\mathbb{R}$ with sufficiently smooth partial derivatives. Given such a function f, we define the functional on $V(\Omega)$ as*

$$L(u) = \int_a^b f(x, u, u')\,dx\,. \tag{3.11}$$

In the next example we establish the link we differential equations. In this example the function f represents the source term in the differential equation, the current in the modeling of fault-current limiter or the burner in the industrial furnace example. The function c and h will turn out later to be the diffusion coefficient and the non-homogeneous term in the Neumann boundary conditions, respectively.

Example 3.1.10. (Two-Dimensional Self-Adjoint Differential Equation) *Given $\Omega \subset \mathbb{R}^2$ an open and bounded two-dimensional domain with boundary $\Gamma = \Gamma_D \cup \Gamma_N$, given c a positive function*

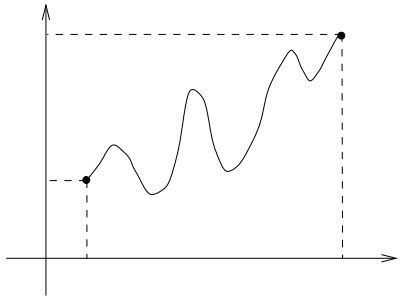


Figure 3.1: *A smooth curve in* $\mathbb{R}^2$ *that connects the points* (a, u_a) *and* (b, u_b)

on Ω, given f a continuous function on Ω and given h a function on Γ_N, a functional on $C^1(\Omega)$ can be defined by

$$L(u) = \int_\Omega \left[1/2\, \nabla u \cdot (c \nabla u) - fu \right] d\Omega + \int_{\Gamma_N} h(x,y)\, u \, d\Gamma \,. \tag{3.12}$$

In turns out in the last example L is linear and bounded (and therefore continuous).

■ 3.1.4 Minimization of Functionals

In this subsection we seek to minimize functionals over the corresponding function space, i.e., to find arguments $\tilde{u} \in V(\Omega)$ that minimize L or

$$\text{find } \tilde{u} \in V(\Omega) \text{ such that } L(\tilde{u}) \leq L(u) \quad \forall u \in V(\Omega) \,. \tag{3.13}$$

Solving this problem is generally hard, if not impossible at all. Local and global minimizers might be hard to distinguish and the global optimum might be non-unique. We therefore aim towards the less ambitious goal of seeking first order critical points of L. These points are defined for functionals in the same way as for functions in several variables. The point $\tilde{u}$ is a first order critical point of L iff the directional derivative of L in $\tilde{u}$ is zero in all possible directions.

For the definition of first order critical point to have sense, it is **important** that the space of possible directions used in the definition forms a vector space, i.e., that this space is closed for linear combinations. This requires the proper treatment of the non-homogeneous Dirichlet boundary conditions. Indeed, any linear combination of two functions u_1 and u_2 satisfying the non-homogeneous Dirichlet boundary conditions no longer satisfies them. This motivates the following definition of the subspace $V_0(\Omega)$ of $V(\Omega)$

$$V_0(\Omega) = \{ v \in V(\Omega) | v|_{\Gamma_D} = 0 \} \,. \tag{3.14}$$

In case that non-homogeneous Dirichlet boundary conditions are imposed the space $V_0(\Omega)$ forms a vector space, while $V(\Omega)$ does not.

The problem of finding (local) minimizers of L now reduces to

$$\text{find } \tilde{u} \in V(\Omega) \text{ such that } \frac{d}{d\epsilon} \big[L(\tilde{u} + \epsilon v) \big]|_{\epsilon=0} = 0 \quad \forall v \in V_0(\Omega) \,. \tag{3.15}$$

Note that $\tilde{u}$ and v not necessarily belong to the same space. We will refer to $V_0(\Omega)$ as the space of test functions. Only if L is convex, there is only one critical point that coincides in the global minimizer of L.

3.2 Calculus of Variations

For future reference we recall here a classical result from variational calculus, which we will refer to as the du Bois-Reymond Lemma (after Paul David Gustav du Bois-Reymond (Germany, 1831 - 1889) (details can be found in e.g. the monographs van Kan et al. [2006]; Strang and Fix [1973]).

Lemma 3.2.1. (One-dimensional du Bois-Reymond's Lemma) *Assume $\Omega = (a,b)$ and let $V_0(\Omega)$ denote a function space of sufficiently smooth function on Ω that vanish on the boundary of*

Ω. *Assume* $f(x)$ *be continuous over* $\overline{\Omega}$, *i.e.,* $f \in C(\Omega)$ *and suppose that*

$$\int_a^b f(x)\, v(x)\, dx = 0, \quad \forall v \in V_0(\Omega)\,. \tag{3.16}$$

Then

$$f(x) = 0 \ \textit{on}\ \Omega. \tag{3.17}$$

The equality (3.19) states that $f(x)$ is zero in *variational* or *weak* form, while (3.20) states that $f(x)$ is zero is strong form. Two functions $f(x)$ and $g(x)$ as said to be equal in weak form if their difference is zero in weak from. The lemma states that for continuous functions the notion on weak and strong form of being zero on interior of the domain Ω coincide. No information on the behavior of $f(x)$ on the boundary of Ω is deduced. In this context the space $V_0(\Omega)$ is called the space of *test functions*. The practical value of this lemma increases as the space $V_0(\Omega)$ can be kept small, in such a way to require to *test* with as few as function $v \in V_0(\Omega)$ as possible. Indeed, if the lemma holds for a given $V_0(\Omega)$ it will automatically be valid for a larger space $Z_0(\Omega)$ (as a subset of test functions $v \in V_0(\Omega) \subset Z_0(\Omega)$ already suffices to draw the conclusion).

Proof. We will prove the lemma for $V_0(\Omega) = C_0^1(\Omega)$ (and therefore by the previous argument also for $V_0(\Omega) = C_0(\Omega)$). We argue by contradiction. Suppose that $f(x_0) > 0$ for any $x_0 \in (a, b)$ (note that the case $f(x_0) < 0$ can be treated similarly) then since $f(x)$ is continuous, there exists a $\delta > 0$ such that

$$f(x) > 0 \quad \text{whenever} \quad |x - x_0| < \delta.$$

Now we choose $v(x) = v_3(x)$ defined by Equation 3.3. Then $v(x) \in C_0^1(\Omega)$, $v(x) > 0$ on $(x_0-\delta, x_0+\delta)$ and

$$\int_a^b f(x)\, v(x)\, dx = \int_{x_0-\delta}^{x_0+\delta} \underbrace{f(x)}_{>0}\, \underbrace{v(x)}_{>0}\, dx > 0. \tag{3.18}$$

This violates the hypothesis (3.19). As x_0 was chosen arbitrarily, we have proven that $f(x) = 0$ on Ω. □

The du Bois-Reymond's Lemma also holds in higher dimensions.

Lemma 3.2.2. (Higher-dimensional du Bois-Reymond's Lemma) *Assume* Ω *to be an open and bounded domain in* $\mathbb{R}^2$ *and let* $V_0(\Omega)$ *denote a function space of sufficiently smooth function on* Ω *that vanish on the boundary of* Ω. *Assume* $f(x, y)$ *be continuous over* $\overline{\Omega}$, *i.e.,* $f \in C(\Omega)$ *and suppose that*

$$\int_\Omega f(x, y)\, v(x, y)\, d\Omega = 0, \quad \forall v \in V_0(\Omega)\,. \tag{3.19}$$

Then

$$f(x, y) = 0 \ \textit{on}\ \Omega. \tag{3.20}$$

3.3 From Functionals to Differential Equations

In this section we illustrate how differential equations can be obtained by the minimizing of functionals. Given a functional L, we will carry out the following three-step procedure:

[1] assume that $\tilde{u}$ is a first-order critical point of L and express that $\tilde{u}$ satisfies the condition (3.15);

[2] integrate integrate-by-parts and exploit the homogeneous Dirichlet boundary conditions of $v \in V_0(\Omega)$;

[3] apply du-Bois Reymond's Lemma to derive the differential equation for $\tilde{u}$ supplied with boundary conditions.

In what follows we illustrate the above procedure using the examples of functionals given above.

3.3.1 Distance in Flat Land

In this subsection we illustrate the above procedure using the functional introduced in Example 3.1.7. Given $\Omega = (a, b)$, and given the function space $V(\Omega)$ and $V_0(\Omega)$ where

$$V(\Omega) = \{u(x) | u''(x) \text{ continuous on } \Omega, u(x = a) = u_a \text{ and } u(x = b) = u_b\} \tag{3.21}$$

and

$$V_0(\Omega) = \{v(x) | v''(x) \text{ continuous on } \Omega, v(x = a) = 0 \text{ and } v(x = b) = 0\}\,, \tag{3.22}$$

we seek to minimize $L(u) = \int_a^b \sqrt{1 + u'(x)^2}\, dx$ over $V(\Omega)$. The motivation for the requirement of $u''(x)$ to be continuous will become clear in the following. We begin by recalling the notation

$$\frac{d}{d\epsilon}(\tilde{u} + \epsilon v) = (\tilde{u} + \epsilon v)' = \tilde{u}' + \epsilon v'\,. \tag{3.23}$$

Next we compute the left-hand side of (3.15)

$$\begin{aligned} \frac{d}{d\epsilon}\big[L(\tilde{u} + \epsilon v)\big] &= \frac{d}{d\epsilon}\Big[\int_a^b \sqrt{1 + (\tilde{u}' + \epsilon v')^2}\, dx\Big] \\ &= \int_a^b \frac{d}{d\epsilon}\sqrt{1 + (\tilde{u}' + \epsilon v')^2}\, dx \\ &= \int_a^b \frac{(\tilde{u}' + \epsilon v')\, v'}{\sqrt{1 + (\tilde{u}' + \epsilon v')^2}}\, dx\,. \end{aligned} \tag{3.24}$$

Therefore

$$\frac{d}{d\epsilon}\big[L(\tilde{u} + \epsilon v)\big]|_{\epsilon=0} = \int_a^b \frac{\tilde{u}'\, v'}{\sqrt{1 + (\tilde{u}')^2}}\, dx\,, \tag{3.25}$$

and using by integration by parts and the homogeneous Dirichlet boundary conditions on v we obtain

$$\frac{d}{d\epsilon}\big[L(\tilde{u} + \epsilon v)\big]|_{\epsilon=0} = \underbrace{\left[\frac{\tilde{u}'\, v}{\sqrt{1 + (\tilde{u}')^2}}\right]_{x=a}^{x=b}}_{=0} - \int_a^b \frac{d}{dx}\left[\frac{\tilde{u}'}{\sqrt{1 + (\tilde{u}')^2}}\right] v\, dx\,. \tag{3.26}$$

We expand the integrand in the right-hand side as

$$\begin{aligned}\frac{d}{dx}\left[\frac{\tilde{u}'}{\sqrt{1+(\tilde{u}')^2}}\right] &= \frac{d}{d\tilde{u}'}\left[\frac{\tilde{u}'}{\sqrt{1+(\tilde{u}')^2}}\right]\frac{d\tilde{u}'}{dx} \\ &= \left[\frac{1}{\sqrt{1+(\tilde{u}')^2}}\right]\tilde{u}'' + \tilde{u}'\frac{d}{d\tilde{u}'}\left[(1+(\tilde{u}')^2)^{-1/2}\right]\tilde{u}'' \\ &= \left[\frac{1+(\tilde{u}')^2}{[1+(\tilde{u}')^2]^{3/2}}\right]\tilde{u}'' - \left[\frac{(\tilde{u}')^2}{[1+(\tilde{u}')^2]^{3/2}}\right]\tilde{u}'' \\ &= \frac{\tilde{u}''}{[1+(\tilde{u}')^2]^{3/2}}\,,\end{aligned} \tag{3.27}$$

and therefore

$$\frac{d}{d\epsilon}\big[L(\tilde{u}+\epsilon v)\big]|_{\epsilon=0} = -\int_a^b \left[\frac{\tilde{u}''}{[1+(\tilde{u}')^2]^{3/2}}\right] v\,dx\,. \tag{3.28}$$

The condition (3.15) that $\tilde{u}$ is a critical point is therefore equivalent to

$$\int_a^b \left[\frac{\tilde{u}''}{[1+(\tilde{u}')^2]^{3/2}}\right] v\,dx = 0 \quad \forall v \in V_0(\Omega)\,. \tag{3.29}$$

If $\tilde{u}(x) \in V(\Omega)$, then $\tilde{u}''(x)$ is continuous and therefore also the function $\tilde{u}''/[1+(\tilde{u}')^2]^{3/2}$ in the integrand in the left-hand side of the above equality. By du-Bois Reymond's Lemma we can then assert that

$$\frac{\tilde{u}''}{[1+(\tilde{u}')^2]^{3/2}} = 0 \text{ on } (a,b) \Leftrightarrow \tilde{u}'' = 0 \text{ on } (a,b)\,, \tag{3.30}$$

which is the differential equation for $\tilde{u}$ sought for.

This example is sufficiently simple to allow for an analytical solution for $\tilde{u}$. Indeed, the differential equation $\tilde{u}'' = 0$ on (a,b) supplied with the boundary conditions $u(x=a) = u_a$ and $u(x=b) = u_b$ yields the solution

$$\tilde{u}(x) = u_a + \frac{u_b - u_a}{b-a}(x-b)\,, \tag{3.31}$$

which is a straight line interval in $\mathbb{R}^2$ between the points (a,u_a) and (b,u_b). More important than the solution is to observe the fact that the differential equation for $\tilde{u}$ can be obtained by minimizing a functional.

■ 3.3.2 Abstract One-Dimensional Minimization Problem

In this subsection we apply the same procedure as before on the functional L described in Example 3.1.9. Given as before $\Omega = (a,b)$, and given the function space $V(\Omega)$ and $V_0(\Omega)$ where

$$V(\Omega) = \{u(x) | u''(x) \text{ continuous on } \Omega, u(x=a) = u_a\} \tag{3.32}$$

and

$$V_0(\Omega) = \{v(x) | v''(x) \text{ continuous on } \Omega, v(x=a) = 0\}\,, \tag{3.33}$$

we seek to minimize $L(u) = \int_a^b f(x,u,u')\,dx$ over $V(\Omega)$. For reasons that will become clear in the following, we require the partial derivative $\partial f/\partial u$ to be continuous and the partial derivative $\partial f/\partial u'$

to be continuously differentiable with respect to x, i.e., we require the derivative $d(\partial f/\partial u')/dx$ to be continuous. Notice the absence of conditions at the end point of Ω in the definition of both $V(\Omega)$ and $V_0(\Omega)$. In our derivation we will also make use of the function space $W_0(\Omega) \subset V_0(\Omega)$ where

$$W_0(\Omega) = \{v(x) | v''(x) \text{ continuous on } \Omega, v(x=a) = 0 \text{ and } v(x=b) = 0\}\,. \tag{3.34}$$

We begin by computing the left-hand side of (3.15)

$$\begin{aligned} \frac{d}{d\epsilon}\big[L(\tilde{u}+\epsilon v)\big] &= \frac{d}{d\epsilon}\big[\int_a^b f(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\,dx\big] \\ &= \int_a^b \frac{d}{d\epsilon} f(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\,dx\,. \end{aligned} \tag{3.35}$$

The derivative in the integrand on the right-hand side can be computed via the chain rule, i.e.,

$$\begin{aligned} \frac{d}{d\epsilon} f(x, \tilde{u}+\epsilon v, \tilde{u}+\epsilon v) &= \frac{\partial f}{\partial u}(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\frac{d(\tilde{u}+\epsilon v)}{d\epsilon} + \\ &\quad \frac{\partial f}{\partial u'}(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\frac{d(\tilde{u}'+\epsilon v')}{d\epsilon} \\ &= \frac{\partial f}{\partial u}(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\,v + \frac{\partial f}{\partial u'}(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\,v'\,. \end{aligned} \tag{3.36}$$

Substituting this result into the integral and splitting the integral into two parts, one obtains

$$\frac{d}{d\epsilon}\big[L(\tilde{u}+\epsilon v)\big] = \int_a^b \frac{\partial f}{\partial u}(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\,v\,dx + \int_a^b \frac{\partial f}{\partial u'}(x, \tilde{u}+\epsilon v, \tilde{u}'+\epsilon v')\,v'\,dx\,. \tag{3.37}$$

Therefore

$$\frac{d}{d\epsilon}\big[L(\tilde{u}+\epsilon v)\big]|_{\epsilon=0} = \int_a^b \frac{\partial f}{\partial u}(x, \tilde{u}, \tilde{u}')\,v\,dx + \int_a^b \frac{\partial f}{\partial u'}(x, \tilde{u}, \tilde{u}')\,v'\,dx\,. \tag{3.38}$$

and using by integration by parts and the homogeneous Dirichlet boundary conditions in $x = a$ on v we obtain

$$\begin{aligned} \frac{d}{d\epsilon}\big[L(\tilde{u}+\epsilon v)\big]|_{\epsilon=0} &= \int_a^b \frac{\partial f}{\partial u}(x, \tilde{u}, \tilde{u}')\,v\,dx + \underbrace{\left[\frac{\partial f}{\partial u'}(x, \tilde{u}, \tilde{u}')\,v\right]_{x=a}^{x=b}}_{=0 \text{ in } x=a \text{ only}} - \int_a^b \frac{d}{dx}\left[\frac{\partial f}{\partial u'}(x, \tilde{u}, \tilde{u}')\right] v\,dx \\ &= \int_a^b \frac{\partial f}{\partial u}(x, \tilde{u}, \tilde{u}')\,v\,dx + \frac{\partial f}{\partial u'}(x=b, \tilde{u}, \tilde{u}')\,v(x=b) - \int_a^b \frac{d}{dx}\left[\frac{\partial f}{\partial u'}(x, \tilde{u}, \tilde{u}')\right] v\,dx\,. \end{aligned} \tag{3.39}$$

The condition (3.15) is then seen to be equivalent to

$$\int_a^b \left[\frac{\partial f}{\partial u} - \frac{d}{dx}\left(\frac{\partial f}{\partial u'}\right)\right] v\,dx = \frac{\partial f}{\partial u'}(x=b, \tilde{u}, \tilde{u}')\,v(x=b) \quad \forall v \in V_0(\Omega)\,. \tag{3.40}$$

The smoothness requirements on the partial derivatives of $f(x, u, u')$ are such that the functions $\partial f/\partial u - d(\partial f/\partial u')/dx$ and $\partial f/\partial u'$ are continuous functions in x allowing us to apply du-Bois Reymond's Lemma. We will in fact apply this lemma twice. The above condition should in particular hold for $v \in W_0(\Omega) \subset V_0(\Omega)$ for which $v(x=b) = 0$ and therefore

$$\int_a^b \left[\frac{\partial f}{\partial u} - \frac{d}{dx}\left(\frac{\partial f}{\partial u'}\right)\right] v\,dx = 0 \quad \forall v \in W_0(\Omega)\,, \tag{3.41}$$

and therefore by du-Bois Reymond's Lemma

$$\boxed{\frac{\partial f}{\partial u} - \frac{d}{dx}\left(\frac{\partial f}{\partial u'}\right) = 0 \text{ on } (a,b)\,.} \tag{3.42}$$

To treat the boundary condition at $x = b$ we return to larger space $V_0(\Omega)$ and to the relation (3.40) knowing now that the left-hand side is zero, i.e.,

$$\frac{\partial f}{\partial u'}(x = b, \tilde{u}, \tilde{u}')\, v(x = b) = 0 \quad \forall v \in V_0(\Omega)\,, \tag{3.43}$$

and therefore we necessarily have that by du-Bois Reymond's Lemma

$$\frac{\partial f}{\partial u'}(x = b, \tilde{u}, \tilde{u}') = 0\,. \tag{3.44}$$

Summarizing, we can state that the solution of the differential equation (3.42) supplied with the boundary conditions $u(x = a) = u_a$ and (3.44) minimizes $L(u) = \int_a^b f(x, u, u')\, dx$ over $V_0(\Omega)$. The minimization of a functional is seen to give raise to a partial differential equation with boundary conditions guaranteeing the latter to have an unique solution.

The equation (3.42) is called the *Euler-Lagrange* equation. It plays an **important** role in various fields of physics such as geometrical optics, classical and quantum mechanics and cosmology as it allows to derive the governing equations starting from a functional describing the total energy of the system. It allows to derive that the shortest path between e.g. two points on the lateral surface of a cylinder is a helix. The Maxwell equations are derived following an energy minimization principle in **?**]. The use of the Euler-Lagrange in deriving the equation of motion for Newton's apple is illustrated in the next example.

Newton's Apple In Example 3.1.8 the function $f(x, u, u')$ is seen to be

$$f(x, u, u') = {}^1\!/_2\, m\, u'(x)^2 - {}^1\!/_2\, k\, u(x)^2\,. \tag{3.45}$$

Therefore

$$\begin{aligned} \frac{\partial f}{\partial u'} &= m\, u'(x) \\ \frac{d}{dx}\left(\frac{\partial f}{\partial u'}\right) &= m\, u''(x) \\ \frac{\partial f}{\partial u} &= -k\, u(x)\,. \end{aligned} \tag{3.46}$$

The Euler-Lagrange equation is therefore seen to simplify

$$m\, u''(x) = -k\, u(x)\,, \tag{3.47}$$

which corresponds to the governing equation for a mass-spring system.

■ 3.3.3 Two-Dimensional Self-Adjoint Problem

In this subsection we seek to minimize the functional $L(u)$ introduced in Example 3.1.10. Given $\Omega \subset \mathbb{R}^2$ an open and bounded two-dimensional domain with boundary $\Gamma = \Gamma_D \cup \Gamma_N$, given c a positive function on Ω, given a continuous function f on Ω, given g a function on Γ_D and h a continuous function on Γ_N, and given the function spaces $V(\Omega)$ and $V_0(\Omega)$ where

$$V(\Omega) = \{u(x,y) | \nabla \cdot (c\nabla u) \text{ continuous over } \Omega, u|_{\Gamma_D} = g\} \tag{3.48}$$

$$V_0(\Omega) = \{v(x,y) | \nabla \cdot (c\nabla v) \text{ continuous over } \Omega, v|_{\Gamma_D} = 0\} \ . \tag{3.49}$$

Notice that the functions in $V_0(\Omega)$ are forced to be zero on $\Gamma_D \subset \Gamma$ only and the imposed smoothness requirements on $u(x,y)$ and $v(x,y)$ coincide with those of the previous one-dimensional example in case that $c(x,y) = 1$. We seek to minimize $L(u)$, where

$$L(u) = \int_\Omega \left[1/2 \, \nabla u \cdot (c\nabla u) - fu \right] d\Omega + \int_{\Gamma_N} h(x,y) \, u \, d\Gamma \tag{3.50}$$

over $V(\Omega)$. In doing so, we will also make use of the space $W_0(\Omega) \subset V_0(\Omega)$ in which the functions are forced to be zero on the whole boundary, i.e.,

$$W_0(\Omega) = \{v(x,y) | \nabla \cdot (c\nabla v) \text{ continuous over } \Omega, v|_{\Gamma} = 0\} \ . \tag{3.51}$$

In this example the left-hand side of (3.15) can be computed as

$$\begin{aligned} L(\tilde{u} + \epsilon v) &= \int_\Omega \left[1/2 \, \nabla(\tilde{u} + \epsilon v) \cdot \left[c\nabla(\tilde{u} + \epsilon v) \right] - f(\tilde{u} + \epsilon v) \right] d\Omega + \\ &\quad \int_{\Gamma_N} h(x,y) \, (\tilde{u} + \epsilon v) \, d\Gamma \\ &= \int_\Omega \left[1/2 \, \nabla \tilde{u} \cdot (c\nabla \tilde{u}) - f\tilde{u} \right] d\Omega + \int_{\Gamma_N} h(x,y) \, \tilde{u} \, d\Gamma + \\ &\quad \epsilon \left[\int_\Omega \left[\nabla v \cdot (c\nabla \tilde{u}) - f \, v \right] d\Omega + \int_{\Gamma_N} h(x,y) \, v \, d\Gamma \right] + \epsilon^2 \int_\Omega 1/2 \, \nabla v \cdot (c\nabla v) \, d\Omega \, . \end{aligned} \tag{3.52}$$

Therefore

$$L(\tilde{u} + \epsilon v)|_{\epsilon=0} = \int_\Omega \left[\nabla v \cdot (c\nabla \tilde{u}) - f \, v \right] d\Omega + \int_{\Gamma_N} h(x,y) \, v \, d\Gamma \tag{3.53}$$

and using integration by parts and the homogeneous boundary conditions of v on Γ_D

$$\begin{aligned} L(\tilde{u} + \epsilon v)|_{\epsilon=0} &= \int_\Omega \nabla v \cdot (c\nabla \tilde{u}) \, v \, d\Omega - \int_\Omega f \, v \, d\Omega + \int_{\Gamma_N} h(x,y) \, v \, d\Gamma \\ &= \int_\Omega \nabla \cdot (c\nabla \tilde{u}) \, v \, d\Omega - \int_\Omega f \, v \, d\Omega + \underbrace{\int_\Gamma c \, \tilde{u}_n \, v \, d\Gamma}_{v=0 \text{ on } \Gamma_D} + \int_{\Gamma_N} h(x,y) \, v \, d\Gamma \\ &= \int_\Omega \left[\nabla \cdot (c\nabla \tilde{u}) - f \right] v \, d\Omega + \int_{\Gamma_N} \left[h(x,y) - c \, \tilde{u}_n \right] v \, d\Gamma \end{aligned} \tag{3.54}$$

The condition (3.15) can thus be equivalently written as

$$\int_\Omega \Big[\nabla\cdot(c\nabla\tilde{u}) - f\Big]\, v\, d\Omega + \int_{\Gamma_N} \big[h(x,y) - c\,\tilde{u}_n\big]\, v\, d\Gamma = 0 \quad \forall v \in V_0(\Omega)\,. \tag{3.55}$$

The imposed smoothness requirements on $u(x,y)$, $f(x,y)$ and $h(x,y)$ are such that we can apply du-Bois Reymond's Lemma. The above condition should in particular hold for $v \in W_0(\Omega) \subset V_0(\Omega)$ for which $v|_\Gamma = 0$ and therefore

$$\int_\Omega \Big[\nabla\cdot(c\nabla\tilde{u}) - f\Big]\, v\, d\Omega = 0 \quad \forall v \in V_0(\Omega)\,, \tag{3.56}$$

and by du-Bois Reymond Lemma's therefore

$$\nabla\cdot(c\nabla\tilde{u}) = f \text{ on } \Omega\,. \tag{3.57}$$

To treat the boundary condition on Γ_N we return to larger space $V_0(\Omega)$ and to the relation (3.55) knowing now that the left-hand side is zero, i.e.,

$$\int_{\Gamma_N} \big[h(x,y) - c\,\tilde{u}_n\big]\, v\, d\Gamma = 0 \quad \forall v \in V_0(\Omega)\,, \tag{3.58}$$

and therefore we necessarily have that by du-Bois Reymond's Lemma

$$c\,\nabla\tilde{u}\cdot\mathbf{n} = c\,\tilde{u}_n = h \text{ on } \ . \tag{3.59}$$

Summarizing we can state the solution of the partial differential (3.57) supplied with the boundary conditions $\tilde{u} = g$ on Γ_D and $c\,\tilde{u}_n = h$ on Γ_N minimizes the functional (3.50) over $V_0(\Omega)$. The minimization of a functional over a function space is again seen to give raise to a partial differential supplied with boundary conditions ensuring the latter to have a unique solution.

3.4 From Differential Equation to Minimization Problem

In the previous section we saw how starting from a minimization problem, a partial differential equation could be derived. In this section we discuss under which conditions a minimization problems can be derived for a given partial differential equation.

The next theorem states that in case that the differential operator is positive and self-adjoint, solving the differential equation is equivalent to minimizing a functional over a function space.

Theorem 3.4.1. *Assume that Ω is an open and bounded domain in $\mathbb{R}$, $\mathbb{R}^2$ or $\mathbb{R}^3$, that $V(\Omega)$ is a space of functions on Ω that are zero on Γ and that f is continuous function on Ω. Suppose that $\mathcal{L}$ is a positive and self-adjoint differential operator on $V(\Omega)$ and that $u_0 \in V(\Omega)$ is the unique solution to the differential equation*

$$\mathcal{L}u = f \tag{3.60}$$

supplied with Dirichlet boundary conditions. Then u_0 minimizes the functional

$$L(u) = \int_\Omega \Big[1/2\, u\,(\mathcal{L}u) - f\, u\Big]\, d\Omega\,, \tag{3.61}$$

meaning that

$$L(u_0) \le L(u) \quad \forall u \in V(\Omega)\,. \tag{3.62}$$

Proof. Consider $u \in V(\Omega)$, $u \neq u_0$. Then as $0 \neq u - u_0 \in V(\Omega)$, we have due to $\mathcal{L}$ being positive on $V(\Omega)$ that

$$1/2 \int_\Omega (u - u_0)\,(\mathcal{L}(u - u_0))\; d\Omega > 0\,. \tag{3.63}$$

Expanding the left-hand side of the previous expression, we obtain

$$\begin{aligned} 1/2 \int_\Omega (u - u_0)\,(\mathcal{L}(u - u_0))\; d\Omega \;\; &= \;\; 1/2 \int_\Omega u\,(\mathcal{L}u)\; d\Omega - 1/2 \int_\Omega u\,(\mathcal{L}u_0)\; d\Omega \\ &\quad -1/2 \int_\Omega u_0\,(\mathcal{L}u)\; d\Omega + 1/2 \int_\Omega u_0\,(\mathcal{L}u_0)\; d\Omega \end{aligned} \tag{3.64}$$

Using the fact that $\mathcal{L}$ is self-adjoint and the $\mathcal{L}u_0 = f$, we have that

$$\begin{aligned} 1/2 \int_\Omega (u - u_0)\,(\mathcal{L}(u - u_0))\; d\Omega \;\; &= \;\; 1/2 \int_\Omega u\,(\mathcal{L}u)\; d\Omega - \int_\Omega u\,(\mathcal{L}u_0)\; d\Omega \\ &\quad -1/2 \int_\Omega u_0\,(\mathcal{L}u_0)\; d\Omega + \int_\Omega f\,u_0\, d\Omega \end{aligned} \tag{3.65}$$

$$\begin{aligned} &= \;\; 1/2 \int_\Omega u\,(\mathcal{L}u)\; d\Omega - \int_\Omega f\,u\, d\Omega \\ &\quad -1/2 \int_\Omega u_0\,(\mathcal{L}u_0)\; d\Omega + \int_\Omega f\,u_0\, d\Omega \end{aligned} \tag{3.66}$$

Given the left-hand side must is positive, we have in fact that

$$\int_\Omega \left[1/2\, u_0\,(\mathcal{L}u_0) - f\,u_0\right] d\Omega \leq \int_\Omega \left[1/2\, u\,(\mathcal{L}u) - f\,u\right] d\Omega \tag{3.67}$$

which is equivalent to what we set out to prove. □

The **importance** of this above theorem resides in the fact that solving the differential equations we are interested in this course (e.g. $\mathcal{L}u = \nabla\cdot(c\nabla u)$) supplied with appropriate boundary conditions to ensure uniqueness is completely equivalent to solving minimizing a functional over a suitable function space. This insight will be useful in treating the Ritz and Galerkin finite elements methods in the next chapters.

4

Variational Formulation and Differential Equations

In the previous chapter we saw the relation between minimization problems and (partial) differential equations. It was demonstrated that if the differential operator is positive and self-adjoint, then, such an associated minimization problem exist. Ritz' finite element method is based on the numerical solution of a minimization problem. To solve problems with differential operators that do not satisfy these requirements, the so-called weak form is introduced. The differential equation is written as a weak form and then a numerical solution to this weak form is determined. This method is more generally applicable and it is the backbone of Galerkin's finite element method.

4.1 Weak forms

Consider the following minimization problem on domain Ω with boundaries $\partial\Omega = \Gamma_1 \cup \Gamma_2$:

$$\textbf{minimization:} \quad \begin{cases} \text{Find } \hat{u} \text{ smooth, such that } \hat{u}|_{\Gamma_1} = g, \text{ and} \\ J(\hat{u}) \leq J(u) \quad \text{for all smooth } u \text{ with } u|_{\Gamma_1} = g, \\ \text{where} \quad J(u) := \frac{1}{2}\int\limits_{\Omega} \|\bar{\nabla} u\|^2 dA. \end{cases} \tag{4.1}$$

Using $u = \hat{u} + \varepsilon v$ for all smooth v with $v|_{\Gamma_1} = 0$, it follows that the solution of equation (4.1) coincides with:

$$\textbf{weak form:} \quad \begin{cases} \text{Find } \hat{u} \text{ smooth, such that } \hat{u}|_{\Gamma_1} = g, \text{ and} \\ \int\limits_{\Omega} \bar{\nabla}\hat{u}.\bar{\nabla} v dA = 0 \quad \text{for all smooth } v \text{ with } v|_{\Gamma_1} = 0. \end{cases} \tag{4.2}$$

It can be demonstrated that the solution of the above exists and that it is unique. We suppose that the boundary of Ω is given by $\Gamma_1 \cup \Gamma_2$. The problems (4.1) and (4.2) have the same solution.

The product rule for differentiation applied to $\int_\Omega \bar{\nabla} u \cdot \bar{\nabla} v dA$ gives:

$$\int_\Omega \bar{\nabla} \cdot [v \bar{\nabla} u] dA - \int_\Omega v \Delta u dA = 0 \quad \forall v \quad \text{with} \quad v|_{\Gamma_1} = 0, \tag{4.3}$$

or

$$\int_{\Gamma_1} v \frac{\partial u}{\partial n} ds + \int_{\Gamma_2} v \frac{\partial u}{\partial n} ds = \int_\Omega v \Delta u dA \quad \forall v|_{\Gamma_1} = 0 \quad (\text{with } v \text{ smooth}), \tag{4.4}$$

with $v|_{\Gamma_1} = 0$, follows

$$\int_{\Gamma_2} v \frac{\partial u}{\partial n} dS = \int_\Omega v \Delta u dA \quad \forall v|_{\Gamma_1} = 0 \quad \text{with } v \text{ smooth}. \tag{4.5}$$

Suppose that $v|_{\Gamma_2} = 0$ besides $v|_{\Gamma_1} = 0$, then du Bois-Reymond Lemma gives $\Delta u = 0$ on Ω. When we do away with the condition $v|_{\Gamma_2} = 0$ and use $\Delta u = 0$ on Ω, we obtain the following natural boundary condition $\frac{\partial u}{\partial n}|_{\Gamma_2} = 0$. Hence smooth solutions of (4.1) and (4.2) satisfy:

$$\textbf{PDE:} \quad \begin{cases} -\Delta u = 0, \\ u|_{\Gamma_1} = g, \\ \frac{\partial u}{\partial n}|_{\Gamma_2} = 0. \end{cases} \tag{4.6}$$

When Δu exists within Ω, then the solutions of equations (4.1), (4.2) and (4.6) are the equal. (4.6) contains a PDE, (4.1) is its corresponding minimization problem and (4.2) is called a *variational formulation* or a *weak form* of PDE (4.6).

So far we went from a weak form to a problem with a PDE. In practice, one often goes the other way around. Since finite element methods are based on either the solution of a minimization problem (such as (4.1)) or a weak form (as in (4.2)), we would like to go from a PDE to a weak from. Further, the condition $v|_{\Gamma_1} = 0$ because of $u|_{\Gamma_1} = g$ and hence prescribed, originates from the use of a minimization problem.

Solving of (4.6) by a numerical solution of the representation of (4.2) is referred to as Galerkin's method. Whereas, aquiring the numerical solution of a representation of (4.1) is called Ritz' method. Galerkin's method is most general: it can always be applied. It doesn't matter whether differential operators are self-adjoint or positive. Therefore, this method will be treated in more detail. The study of minimization problems was needed to motivate the condition $v = 0$ on locations where u is prescribed (by an essential condition).

A major advantage of the weak form (4.2) is the fact it is easier to prove existence and uniqueness for (4.2) than for (4.6). It is clear that a solution of (4.6) always is always a solution of (4.2). A solution of the PDE (4.6) always needs the second order derivatives to exist, whereas in the solution of (4.2) only the integrals have to exist. For the solutions of the weak form, it may be possible that the second order derivatives do not exist at all. For that reason, the term *weak form* or *weak solution* is used for the problem and its solution respectively.

The function v is commonly referred to as a 'test function'. Let's go from (4.6) to (4.2). Given $\Delta u = 0 \quad \Leftrightarrow \quad v \Delta u = 0$ for all $v|_{\Gamma_1} = 0$ (reason is that $u|_{\Gamma_1} = g$ is prescribed!) then

$$\int_\Omega v\Delta u dA = 0, \tag{4.7}$$

$$\int_\Omega \bar{\nabla}\cdot[v\bar{\nabla}u]dA - \int_\Omega \bar{\nabla}u\cdot\bar{\nabla}vdA = 0 \quad \forall v|_{\Gamma_1} = 0. \tag{4.8}$$

The product rule for differentiation was used here. Using the Divergence Theorem, this gives (since $v|_{\Gamma_1} = 0$)

$$\int_{\Gamma_2} v\frac{\partial u}{\partial n}ds - \int_\Omega \bar{\nabla}u\cdot\bar{\nabla}vdA = 0 \quad \forall \quad v|_{\Gamma_1} = 0. \tag{4.9}$$

Since in (4.6) it is required that $\frac{\partial u}{\partial n}|_{\Gamma_2} = 0$, we obtain

$$\int_{\Gamma_2} v\frac{\partial u}{\partial n}ds = 0, \tag{4.10}$$

and hence,

$$\int_\Omega \bar{\nabla}u\cdot\bar{\nabla}vdA = 0 \quad \forall \quad v|_{\Gamma_1} = 0 \quad \text{smooth}. \tag{4.11}$$

Hence (4.6) is equivalent to (4.2), if we are not bothered by the smoothness considerations:

$$\begin{cases} \text{Find } \hat{u} \text{ smooth, such that } \hat{u}|_{\Gamma_1} = g, \text{ and} \\ \int_\Omega \bar{\nabla}u\cdot\bar{\nabla}vdA = 0 \quad \text{for all smooth } v \text{ with } v|_{\Gamma_1} = 0. \end{cases} \tag{4.12}$$

Here (4.2) is also sometimes referred to as the finite element formulation of (4.6). The same principle may be applied to

$$\begin{cases} \frac{\partial u}{\partial t} = \Delta u + f, \\ u|_{\Gamma_1} = g, \\ \frac{\partial u}{\partial n}|_{\Gamma_2} = h, \\ u(x,y,0) = 0, \quad t = 0 \quad (x,y) \in \Omega. \end{cases} \tag{4.13}$$

In the above problem the boundary of the domain of computation Ω is given by $\Gamma_1 \cup \Gamma_2$. The question is now to find a finite element formulation for (4.13). We multiply the PDE with a testfunction v, that satisfies $v|_{\Gamma_1} = 0$, since $u|_{\Gamma_1} = g$ is prescribed, to obtain, after integration over Ω,

$$\int_\Omega \frac{\partial u}{\partial t}vdA = \int_\Omega v\Delta udA + \int_\Omega fvdA \quad \forall v\ |_{\Gamma_1} = 0, \tag{4.14}$$

(v smooth). Using the product rule for differentiation, we obtain

$$\int_\Omega \frac{\partial u}{\partial t}vdA = \int_\Omega \bar{\nabla}\cdot[v\bar{\nabla}u]dA - \int_\Omega \bar{\nabla}u\cdot\bar{\nabla}vdA + \int_\Omega fvdA, \quad \forall v\ |_{\Gamma_1} = 0. \tag{4.15}$$

The Divergence Theorem implies:

$$\int_\Omega \frac{\partial u}{\partial t} v dA = \int_{\Gamma_1} \frac{\partial u}{\partial n} v ds + \int_{\Gamma_2} \frac{\partial u}{\partial n} v ds - \int_\Omega \bar{\nabla} u \cdot \bar{\nabla} v dA + \int_\Omega f v dA, \quad \forall v \mid_{\Gamma_1} = 0. \tag{4.16}$$

Since $\dfrac{\partial u}{\partial n} = h$ on Γ_2 and $v \mid_{\Gamma_1} = 0$, we obtain: Find u with $u \mid_{t=0} = 0$, $u \mid_{\Gamma_1} = g$ such that

$$\int_\Omega \frac{\partial u}{\partial t} v dA = \int_{\Gamma_2} h v ds - \int_\Omega \bar{\nabla} u \cdot \bar{\nabla} v dA + \int_\Omega f v dA, \quad \forall v \mid_{\Gamma_1} = 0. \tag{4.17}$$

Equation (4.17) is the variational form or finite element form of (4.13). Note that the Neumann BC is changed into a line-integral over Γ_2. Of course, it is easy to show that (4.13) can be derived, once only (4.17) is given:

$$\int_\Omega \left[\frac{\partial u}{\partial t} - f \right] v dA = \int_{\Gamma_2} h v ds - \int_\Omega \bar{\nabla} \cdot [v \bar{\nabla} u] dA + \int_\Omega v \Delta u dA, \quad \forall v \mid_{\Gamma_1} = 0. \tag{4.18}$$

Using $v \mid_{\Gamma_1} = 0$, this gives

$$\int_\Omega \left[\frac{\partial u}{\partial t} - \Delta u - f \right] v dA = \int_{\Gamma_2} \left[h - \frac{\partial u}{\partial n} \right] v ds \quad \forall v \mid_{\Gamma_1} = 0. \tag{4.19}$$

If we set $v \mid_{\Gamma_2} = 0$ besides $v \mid_{\Gamma_1} = 0$, we obtain from du Bois-Reymond

$$\int_\Omega \left[\frac{\partial u}{\partial t} - \Delta c - f \right] v dA = 0 \Rightarrow \frac{\partial u}{\partial t} - \Delta u - f = 0 \quad \text{on } \Omega. \tag{4.20}$$

This implies after releasing $v \mid_{\Gamma_2} = 0$ that $h - \dfrac{\partial u}{\partial n} = 0$ on Γ_2 (again from du Bois-Reymond). We see that (4.17) corresponds with (4.13), since we require $u \mid_{\Gamma_1} = g$ and $u \mid_{t=0} = 0$ for both (4.13) and (4.17). Note that for the derivation of the weak form, we always multiply the PDE with a test-function v, which must satisfy $v = 0$ on a boundary with a Dirichlet condition (*i.e.* an essential condition). Subsequently we integrate over the domain of computation.

Exercise 4.1.1. *Suppose that we have been given the following problem:*

$$\begin{cases} \dfrac{\partial u}{\partial t} = \Delta u & \text{on } \Omega, \\ u\,|_{\Gamma_1} = g & \text{on } \Gamma_1, \\ u\,|_{\Gamma_2} + \dfrac{\partial u}{\partial n}\,|_{\Gamma_2} = h & \text{on } \Gamma_2, \\ u(x,y,0) = 0 & \text{for } t = 0 \text{ on } \Omega. \end{cases} \tag{4.21}$$

Show that a weak form of the above problem is given by:
Find u smooth, subject to $u\,|_{\Gamma_1} = g$ and $u\,|_{t=0} = 0$, such that

$$\int_\Omega \frac{\partial u}{\partial t} v dA = -\int_\Omega \bar{\nabla} u \cdot \bar{\nabla} v dA + \int_{\Gamma_2} (h-u)\, v ds \quad \forall v\,|_{\Gamma_1} = 0. \tag{4.22}$$

v smooth.
The above weak form (4.22) is used to solve (4.21) by the use of finite elements. Note that the Robin-condition is a natural boundary condition, which is contained in the weak form in the second term of the right-hand of the (4.22).

4.2 Which weak formulation?

When we considered

$$\begin{cases} \Delta u = f & \text{on } \Omega, \\ u\,|_\Gamma = 0, \end{cases} \tag{4.23}$$

then we saw that a weak form is given by (**??**)

$$\begin{cases} \text{Find } u\,|_\Gamma = 0 \text{ such that,} \\ -\displaystyle\int_\Omega \nabla u \cdot \nabla v dA = \int_\Omega f v dA \quad \text{for all } v\,|_\Gamma = 0. \end{cases} \tag{4.24}$$

The above problem is a weak form, but the following problem is also a weak form:

$$\begin{cases} \text{Find } u\,|_\Gamma = 0 \text{ such that,} \\ \displaystyle\int_\Omega v \Delta u dA = \int_\Omega f v dA \quad \text{for all } v\,|_\Gamma = 0, \end{cases} \tag{4.25}$$

or even

$$\begin{cases} \text{Find } u\,|_\Gamma = 0 \text{ such that,} \\ -\displaystyle\int_\Omega u \Delta v dA = \int_\Omega f v dA \quad \text{for all } v\,|_\Gamma = 0. \end{cases} \tag{4.26}$$

Forms (4.24),(4.25) and (4.26) are all possible weak forms of (4.23). However, in the finite element calculations, (4.25) and (4.26) are not common. This is due to the reduction of order of the derivatives in the first form (4.24). Here only the first derivatives are used and this will give an

advantage for the implementation of the FEM, which we will see later. A more important advantage is that for a minimized order of derivatives in the weak form, the class of allowable solutions is largest, in the sense that solutions that are less smooth are allowable. As a rule of thumb, now, we mention that when we derive a weak form, then we should try to minimize the highest order of the derivatives that occur in the integrals.

Example:

$$\begin{cases} u'''' = f \quad f \text{ on } x \in (0,1), \\ u(0) = 0, \\ u'(0) = 0, \\ u(1) = 0, \\ u'(1) = 0. \end{cases} \tag{4.27}$$

We derive a weak form with the lowest order for the derivatives.

$$\int_0^1 u''''v dx = \int_0^1 fv dx \quad \text{for all } v(0) = 0 = v'(0) = v(1) = v'(1), \tag{4.28}$$

partial integration gives

$$\left[u'''v\right]_0^1 - \int_0^1 u'''v' dx = \int_0^1 fv dx \quad \text{for all } v(0) = 0 = v'(0) = v(1) = v'(1), \tag{4.29}$$

with the condition $v(0) = 0 = v(1)$ follows

$$-\int_0^1 u'''v' dx = \int_0^1 fv dx \quad \text{for all } v(0) = 0 = v'(0) = v(1) = v'(1). \tag{4.30}$$

Partial integration, again, gives (with $v'(1) = 0 = v'(0)$)

$$\int_0^1 u''v'' dx = \int_0^1 fv dx \quad \text{for all } v(0) = 0 = v'(0) = v(1) = v'(1). \tag{4.31}$$

Now we stop, because, another partial integration would increase the maximum order of the derivatives again to obtain

$$\int_0^1 u'v''' dx = \int_0^1 fv dx \quad \text{for all } v(0) = 0 = v'(0) = v(1) = v'(1). \tag{4.32}$$

We do not use (4.32) but (4.31) as the weak form for the finite element method.

4.3 Mathematical considerations:*existence and uniqueness*

This section is intended for the interested reader and it is not necessary for the understanding of the implementation of the finite element method. Consider the following Poisson problem

$$\begin{aligned} -\Delta u &= f(\mathbf{x}), \qquad \text{for } \mathbf{x} \in \Omega, \\ u &= g(\mathbf{x}), \qquad \text{for } \mathbf{x} \in \partial\Omega. \end{aligned} \tag{4.33}$$

Here we assume that $f(\mathbf{x})$ and $g(\mathbf{x})$ are given continuous functions. Let $\overline{\Omega} = \Omega \cup \partial\Omega$ be the closure of Ω, then the following assertion can be demonstrated:

Theorem 4.3.1. *If $f(\mathbf{x})$ and $g(\mathbf{x})$ are continuous and if the boundary curve is (piecewise) smooth, then problem* (4.33) *has one and only one solution such that $u \in C^2(\Omega) \cap C^1(\overline{\Omega})$ (that is the solution has continuous partial derivatives up to at least the second order over the open domain Ω and at least continuous first order partial derivatives on the boundary).*

We will not prove this result for the existence and uniqueness of a classical solution to problem (4.33). The proof of the above theorem is far from trivial, the interested reader is referred to the monograph Evans [1999] for instance. The fact that the second order partial derivatives need to be continuous is a rather strong requirement.

The finite element representation of the above problem is given by

$$\begin{aligned} &\text{Find } u \in H^1(\Omega), \text{ subject to } u = g \text{ on } \partial\Omega, \text{ such that} \\ &\int_\Omega \nabla u \cdot \nabla \phi d\Omega = \int_\Omega \phi f d\Omega, \text{ for all } \phi \in H^1(\Omega). \end{aligned} \tag{4.34}$$

In the above problem, the notation $H^1(\Omega)$ has been used, this concerns the set of functions for which the integral over Ω of the square of the function and its gradient is finite. Informally speaking, this is

$$u \in H^1(\Omega) \Longleftrightarrow \int_\Omega u^2 d\Omega < \infty \text{ and } \int_\Omega ||\nabla u||^2 d\Omega < \infty. \tag{4.35}$$

This set of functions represents a *Hilbert space* and is commonly referred to as a *Sobolev space*. Using the fact that each function that is in $H^1(\Omega)$ is also continuous on Ω, that is $H^1(\Omega) \subset C^0(\Omega)$, the following claim can be proved

Theorem 4.3.2. *The problem* (5.86) *has one and only one solution u, such that $u \in H^1(\Omega)$.*

The proof of the above theorem resides on the Lax-Milgram Theorem (see for instance the book by Kreyszig [1989]):

Theorem 4.3.3. *Let V be a Hilbert space and $a(\cdot,\cdot)$ a bilinear form on V, which is*

[1] *bounded: $|a(u,v)| \leq C\|u\|\|v\|$ and*

[2] *coercive: $a(u,u) \geq c\|u\|^2$.*

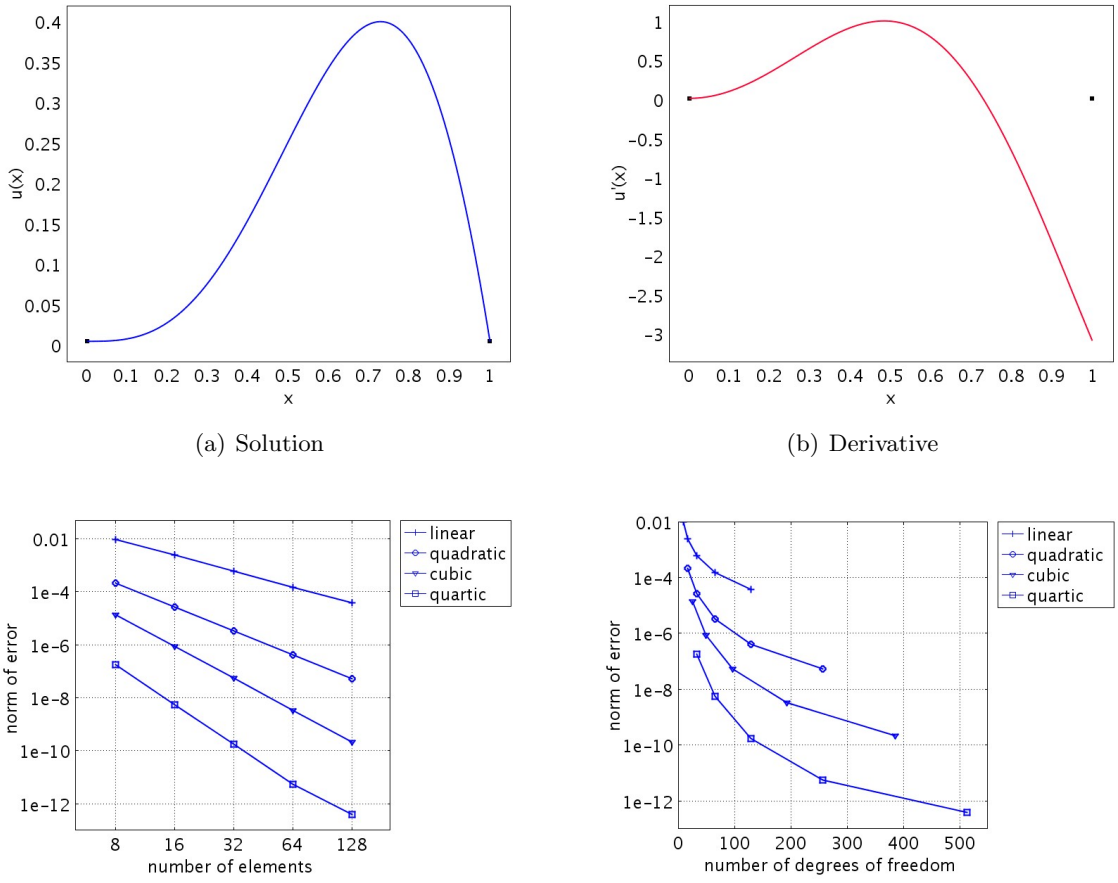
Then, for any linear bounded functional $f \in V'$, there is a unique solution $u \in V$ to the equation

$$a(u,v) = f(v), \text{ for all } v \in V. \tag{4.36}$$

The proof takes into account the fact that the linear operator is positive (more exactly speaking coercive, which is $\int_\Omega ||\nabla u||^2 d\Omega \geq \alpha \int_\Omega u^2 d\Omega$ for some $\alpha > 0$) and continuous. Further, the right hand side represents a bounded linear functional. These issues constitute the hypotheses under which the Lax-Milgram theorem holds and hence have to be demonstrated. In this text the (straighforward) proof is omitted and a full proof of the above theorem can be found in van Kan et al. [2006] for instance.

The most important lesson that we learn here, is that the solution to the weak form exists and that it is uniquely defined. Further, the weak form allows a larger class of functions as solutions than the PDE does.

Smooth Solution In this paragraph we study the convergence of the FEM method in case that the solution and its derivative are smooth. As exact solution we use the function $u(x) = x^2 \sin(\pi x)$ on the interval $0 \leq x \leq 1$. Figure **??** shows the theoretically predicted convergence behavior.



(a) Solution

(b) Derivative

(c) Accurarcy versus mesh size

(d) Accuracy versus problem size

Figure 4.1: *Convergence study for a smooth problem.*

Solution with Discontinuous Derivative In paragraph we study the convergence of the FEM method in case that diffusion coefficient $c(x)$ has a jump-discontinuity. We take $\Omega = (0, 2)$, and set

$$c(x) = \begin{cases} c_1 \text{ if } 0 \leq x \leq 1 \\ c_2 \text{ if } 1 < x \leq 2 \end{cases}, \tag{4.37}$$

and

$$u(x) = \begin{cases} \frac{1}{42}x^7 - \frac{1}{42(c_1+c_2)}(7\,c_1 + 79\,c_2)x \text{ if } 0 \leq x \leq 1 \\ \frac{1}{42}x^7 - \frac{1}{42(c_1+c_2)}(79\,c_1 + 7\,c_2)x + \frac{12}{7}\frac{c_1-c_2}{c_1+c_2} \text{ if } 1 < x \leq 2 \end{cases}, \tag{4.38}$$

5

Galerkin's Finite Element Method

In this chapter we treat the finite element method, which was proposed by Galerkin. The method is based on the weak form of the PDE. This Galerkin method is more general than the Ritz' method, which is based on the solution of a minimization problem and, hence, is only suitable whenever the differential operator is positive and self-adjoint.

5.1 The principle of Galerkin's method

Given the following weak form:

$$\begin{cases} \text{Find } u \mid_{\Gamma} = 0 \text{ such that,} \\ \int\limits_{\Omega} \bar{\nabla} u \cdot \bar{\nabla} v \, dA = \int\limits_{\Omega} f \, v \, dA \quad \text{for all } v \mid_{\Gamma} = 0. \end{cases} \tag{5.1}$$

Here Ω is a general simply connected domain in $\mathbb{R}^1$ or $\mathbb{R}^2$ or $\mathbb{R}^3$. A crucial principle for the FEM is that we write u as a sum of *basis-functions* $\varphi_i(x, y)$, which satisfy

$$\varphi_i(x, y) \mid_{\Gamma} = 0, \tag{5.2}$$

i.e. hence

$$u(x, y) = \sum_{j=1}^{\infty} c_j \varphi_j(x, y). \tag{5.3}$$

Since, we cannot do calculations with an infinite number of terms, we truncate this series such that we only take the first n terms into account, then

$$\hat{u}(x, y) = \sum_{j=1}^{n} c_j \varphi_j(x, y), \tag{5.4}$$

where $\hat{u}$ denotes the approximation of the solution of (5.1). As an example for φ_i one might take powers, sines, cosines (viz. Fourier) and so on. We will assume here that

$$\hat{u}(x,y) = \sum_{j=1}^{n} c_j \varphi_j(x,y) \to u(x,y) \quad \text{as } n \to \infty, \tag{5.5}$$

note that $\hat{u}(x,y)$ represents the approximated solution of (5.1) and $u(x,y)$ the exact solution of (5.1) respectively. There is a lot of mathematical theory needed to prove that $\hat{u} \to u$ as $n \to \infty$ for a specific set of basis-functions $\varphi_i(x,y)$. The Finite Element representation of weak form (5.1) is:

$$\begin{cases} \text{Find the set of constants } \{c_1, \ldots, c_n\} \text{ such that,} \\ \displaystyle\sum_{j=1}^{n} \int_{\Omega} c_j \bar{\nabla}\varphi_j(x,y) \cdot \bar{\nabla}\varphi_i(x,y)\, dA = \int_{\Omega} f\, \varphi_i\, dA \quad \text{for all } i \in \{1, \ldots, n\}. \end{cases} \tag{5.6}$$

Note that we assume here that all functions $v \mid_{\Gamma} = 0$ are represented by (linear combinations of) the set $\varphi_i(x,y)$, $i \in \{1, \ldots, n\}$. We will use this assumption and skip the mathematical proof (see Strang and Fix [1973], Cuvelier et al. [1986] and Braess [1996] for instance for a proof). Further, in (5.6), we will make a choice for the functions $\{\varphi_i(x,y)\}$ and hence they are known in the Finite Element calculations. It turns out that the choice of the basis-functions $\{\varphi_i(x,y)\}$ influences the accuracy and speed of computations. The accuracy is a difficult subject, which we will treat without detail. The speed of computation is easier to deal with. Note that $\bar{u} \mid_{\Gamma} = 0$ due to $\varphi_i(x,y) \mid_{\Gamma} = 0$, $i \in \{1, \ldots, n\}$. (5.6) implies a set of linear equations of $\{c_i\}$:

$$\begin{array}{c} c_1 \int_{\Omega} \bar{\nabla}\varphi_1 \cdot \bar{\nabla}\varphi_1 dA + c_2 \int_{\Omega} \bar{\nabla}\varphi_2 \cdot \bar{\nabla}\varphi_1 dA + c_3 \int_{\Omega} \bar{\nabla}\varphi_3 \cdot \bar{\nabla}\varphi_1 dA + \cdots + c_n \int_{\Omega} \bar{\nabla}\varphi_n \cdot \bar{\nabla}\varphi_1 dA = \int_{\Omega} f\, \varphi_1\, dA, \\ c_1 \int_{\Omega} \bar{\nabla}\varphi_1 \cdot \bar{\nabla}\varphi_2 dA + c_2 \int_{\Omega} \bar{\nabla}\varphi_2 \cdot \bar{\nabla}\varphi_2 dA + c_3 \int_{\Omega} \bar{\nabla}\varphi_3 \cdot \bar{\nabla}\varphi_2 dA + \cdots + c_n \int_{\Omega} \bar{\nabla}\varphi_n \cdot \bar{\nabla}\varphi_2 dA = \int_{\Omega} f\, \varphi_2\, dA \\ \vdots \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \vdots \\ c_1 \int_{\Omega} \bar{\nabla}\varphi_1 \cdot \bar{\nabla}\varphi_n dA + c_2 \int_{\Omega} \bar{\nabla}\varphi_2 \cdot \bar{\nabla}\varphi_n dA + c_3 \int_{\Omega} \bar{\nabla}\varphi_3 \cdot \bar{\nabla}\varphi_n dA + \cdots + c_n \int_{\Omega} \bar{\nabla}\varphi_n \cdot \bar{\nabla}\varphi_n dA = \int_{\Omega} f\, \varphi_n\, dA, \end{array} \tag{5.7}$$

The *discretization* matrix here is refered to as the stiffness matrix, its elements are

$$A_{ij} = \int_{\Omega} \bar{\nabla}\varphi_i \cdot \bar{\nabla}\varphi_j dA \tag{5.8}$$

Exercise 5.1.1. *Show that A is symmetric,* i.e. $a_{ij} = a_{ji}$.

For a fast solution of the system of linear equations, one would like A to be as sparse as possible (*i.e.* A should contain as many zeros as possible). When $\{\varphi_i\}$ are orthogonal over the $\bar{\nabla}$, then

$$\int_{\Omega} \bar{\nabla}\varphi_i \cdot \bar{\nabla}\varphi_j dA = 0 \quad \text{when } i \neq j. \tag{5.9}$$

This would be ideal: the c_j then follows very easily from solving a linear system with diagonal coefficient matrix:

$$c_j = \frac{\int_{\Omega} f\, \varphi_j\, dA}{\int_{\Omega} \nabla\varphi_j \nabla\varphi_j\, dA}, \cdot \tag{5.10}$$

In practice it is not always possible to choose a set of orthogonal basis-functions, but we try to choose a set that is almost orthogonal. This means that A consists of zeroes mainly (*i.e.* A is a sparse matrix). We will choose basis-functions $\{\varphi_i(x, y)\}$ that are piecewise linear. Suppose that the domain of computation is divided into a set of gridnodes, see below, with numbers for the unknowns (figure 5.1).

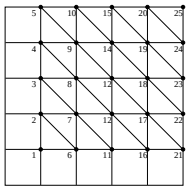


Figure 5.1: *An example of a domain divided into a Finite Element mesh*

Then, we will choose $\varphi_i(x, y)$ to be piecewise (bi-)linear, such that

$$\varphi_i(x_j, y_j) = \begin{cases} 1, & \text{for } (x_j, y_j) = (x_i, y_i), \\ 0, & \text{for } (x_j, y_j) \neq (x_i, y_i). \end{cases} \tag{5.11}$$

The reason for this choice will be motivated by the use of a one-dimensional example. It is clear that for this case the integrals of,

$$\int_{\Omega} \nabla\varphi_i \cdot \nabla\varphi_j dA, \tag{5.12}$$

only do not vanish when i and j are equal or when i and j are neighboring gridpoints, due to piecewise linearity of the basis-functions $\{\varphi_i\}$. This implies that the basis-functions have a compact support and hence, the stiffness-matrix will be sparse. First we motivate the choice of piecewise linear basis functions by the use of a one dimensional example.

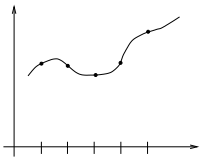


Figure 5.2: *The function $u(x)$ at gridnodes $\{x_i\}$*

5.2 Motivation of piecewise linear basis-functions

Given any function $u = u(x)$ and a division of gridnodes, see figure 5.2. For $[x_{i-1}, x_i]$, we approximate $u(x)$ by the use of linear interpolation:

$$\bar{u}(x) = u(x_{i-1}) + \frac{u(x_i) - u(x_{i-1})}{x_i - x_{i-1}}(x - x_{i-1}) \quad \text{for } x \in [x_{i-1}, x_i], \tag{5.13}$$

or with $u_i = u(x_i)$, we obtain (figure 5.3)

$$\begin{aligned} \bar{u}(x) &= u_{i-1} + \frac{u_i - u_{i-1}}{x_i - x_{i-1}}(x - x_{i-1}), \\ &= u_{i-1}\left\{1 + \frac{x_{i-1} - x}{x_i - x_{i-1}}\right\} + u_i\left\{\frac{x - x_{i-1}}{x_i - x_{i-1}}\right\}, \\ &= u_{i-1}\frac{x_i - x}{x_i - x_{i-1}} + u_i\frac{x - x_{i-1}}{x_i - x_{i-1}} =: u_{i-1}l_{i-1}(x) + u_i l_i(x). \end{aligned} \tag{5.14}$$

Hence

$$\bar{u}(x) = u_{i-1}l_{i-1}(x) + u_i l_i(x) \quad \text{for } x \in [x_{i-1}, x_i]. \tag{5.15}$$

We do the same for $x \in [x_i, x_{i+1}]$ (see Figure 5.3), to obtain:

$$u(x) = u_i l_i(x) + u_{i+1}l_{i+1}(x) \quad \text{on } x \in [x_i, x_{i+1}]. \tag{5.16}$$

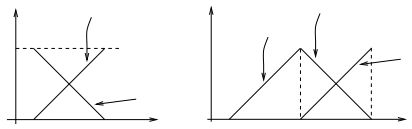


Figure 5.3: *The piecewise linear functions* $\{l_i(x)\}$

Therewith, we write for the approximation $\bar{u}(x)$ of $u(x)$:

$$\bar{u}(x) = \sum_{j=1}^{n} u_j \phi_j(x), \tag{5.17}$$

where

$$\phi_i(x) = \begin{cases} \dfrac{x - x_{i-1}}{x_i - x_{i-1}}, & \text{for } x \in [x_{i-1}, x_i], \\ \dfrac{x - x_{i+1}}{x_i - x_{i+1}}, & \text{for } x \in [x_i, x_{i+1}], \\ 0, & \text{for } x \notin [x_{i-1}, x_{i+1}]. \end{cases} \tag{5.18}$$

Hence $\phi_i(x)$ is piecewise linear, see Figure 5.4, where

$$\phi_i(x_j) = \begin{cases} 0, & i \neq j, \\ 1, & i = j. \end{cases} \tag{5.19}$$

In the finite element method we put in the functions $\{\phi_i(x)\}$ and we determine $\{u_i\}$. Therewith, we obtain the solution.

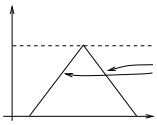


Figure 5.4: *The piecewise linear function $\phi_i(x)$*

5.3 Evaluation of a one-dimensional example

Now we treat procedures to approximate the integrals. For this set of basis-functions we solve: Find $u(x)$ such that $u(0) = 0$ and

$$-\int_0^1 u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx, \quad \text{for all } v(0) = 0. \tag{5.20}$$

Exercise 5.3.1. *Find the differential equation for u with all boundary conditions that corresponds to (5.20) with $u(0) = 0$.*

The approximation $\hat{u}(x) = \sum_{j=1}^{n} u_j\phi_j(x)$ (where $\hat{u} \to u$ as $n \to \infty$ is assumed) leads to the following Finite Element formation:

Find coefficient $\{u_i\}$, in $\hat{u}(x) = \sum_{j=1}^{n} u_j\phi_j(x)$, with

$$\phi_i(x_j) = \begin{cases} 0, & i \neq j, \\ 1, & i = j, \end{cases} \tag{5.21}$$

such that

$$-\int_0^1 \sum_{j=1}^{n} u_j \phi_j'(x)\phi_i'(x)dx = \int_0^1 f(x)\phi_i(x)dx \quad \forall i \in \{1,\ldots,n\}. \tag{5.22}$$

The computation of $\int_0^1 \phi_i'(x)\phi_j'(x)dx$ is simple, since $\{\phi_i\}$ are piecewise linear functions

$$\int_0^1 \phi_{i-1}'(x)\phi_i'(x)dx = \int_{x_{i-1}}^{x_i} \frac{1}{x_{i-1}-x_i}\frac{1}{x_i-x_{i-1}}dx = -\frac{1}{x_i-x_{i-1}}, \tag{5.23}$$

$$\int_0^1 \phi_i'(x)\phi_i'(x)dx = \frac{1}{x_i-x_{i-1}} + \frac{1}{x_{i+1}-x_i}, \tag{5.24}$$

$$\int_0^1 \phi_i'(x)\phi_{i+2}'(x)dx = 0, \tag{5.25}$$

(why?). Of course, given any function $f(x)$ whose product with $\phi_i(x)$ is integrable, the integral $\int_0^1 f(x)\phi_i(x)dx$ may be evaluated. For some cases it is even possible to find an anti-derivative for $\phi_i(x)f(x)$. However, for many cases this isn't possible. We then evaluate the integral numerically. Standard FE-packages use the Newton-Cotes formulas, which read as

$$\begin{aligned}\int_0^1 f(x)\phi_i(x)dx &\approx f(x_i)\int_0^1 \phi_i(x)dx \\ &= f(x_i)\int_{x_{i-1}}^{x_{i+1}} \phi_i(x)dx \\ &= f(x_i)\frac{1}{2}(x_{i+1}-x_{i-1}) \quad \text{for } i \in \{1,2,\ldots,n-1\}.\end{aligned} \tag{5.26}$$

For $i = n$ we use

$$\int_0^1 f(x)\phi_n(x)dx \approx \int_{x_{n-1}}^{x_n} f(x_n)\phi_n(x)dx = f(x_n)\frac{1}{2}(x_n - x_{n-1}), \tag{5.27}$$

or better

$$\frac{1}{2} f\left(\frac{x_n + x_{n-1}}{2}\right)(x_n - x_{n-1}). \tag{5.28}$$

Evaluation of all integrals in (5.22) gives a system of linear equations in $\{u_i\}$. We also know that $u'(1) = 0$ results from (5.20) as a natural boundary condition. This condition should also follow

from (5.22):

$$\int_0^1 \phi'_{n-1}\phi'_n dx = -\frac{1}{x_n - x_{n-1}}, \tag{5.29}$$

$$\int_0^1 \phi'_n\phi'_n dx = \frac{1}{x_n - x_{n-1}}, \tag{5.30}$$

$$\Rightarrow -\frac{1}{x_n - x_{n-1}}u_{n-1} + \frac{1}{x_n - x_{n-1}}u_n = f(\bar{x}_n)\frac{1}{2}(x_n - x_{n-1}), \tag{5.31}$$

where $\bar{x}_n = \dfrac{x_n + x_{n-1}}{2}$ can be chosen (or just $\bar{x}_n = x_n$). Further $f(\bar{x}_n) = u''(\bar{x}_n)$ (see exercise 5.3.1). Implies

$$\begin{aligned} \frac{u_n - u_{n-1}}{x_n - x_{n-1}} &= \frac{1}{2}f(\bar{x}_n)(x_n - x_{n-1}) \\ &= \frac{1}{2}u''(\bar{x}_n)(x_n - x_{n-1}). \end{aligned} \tag{5.32}$$

This implies that $\dfrac{u_n - u_{n-1}}{x_n - x_{n-1}} \to 0$ as $n \to \infty$ ($x_n - x_{n-1} \to 0$). Hence the natural boundary condition is recovered.

5.4 Ritz' method of finite elements by a simple example

Suppose that we have the following minimization problem:

$$\min_{u \in U} \quad J(u), \text{with} \quad J(u) = \frac{1}{2}\int_0^1 \left(\frac{du}{dx}\right)^2 dx, \tag{5.33}$$

where $U := \{u \text{ smooth: } u(0) = 0\}$. Then this problem corresponds to the solution of

$$\begin{cases} \dfrac{d^2u}{dx^2} = 0 \quad \text{on } (0,1), \\ u(0) = 0, \\ \dfrac{du}{dx}(1) = 0. \end{cases} \tag{5.34}$$

Note that the solution is given by $u = 0$. Now we use Ritz' FEM to solve the minimization problem. Let the approximate solution be given by (Figure 5.4)

$$\hat{u}_n(x) = \sum_{j=1}^{n} u_j\phi_j(x), \tag{5.35}$$

where we assume that $\hat{u}_n(x) \to u(x)$ on $x \in [0,1]$ as $n \to \infty$. Then we look for constants $\{u_1, u_2, \ldots, u_n\}$ such that $J(u)$ is minimal. In other words:

$$\frac{\partial}{\partial u_i}J(u) = 0 \quad \forall i \in \{1, 2, 3, \ldots, n\}. \tag{5.36}$$

Substitution of $\sum_{j=1}^{n} u_j\phi_j(x) = \bar{u}_n(x)$ into $J(u)$ gives:

$$J(u) = \frac{1}{2}\int_0^1 \sum_{i=1}^{n} u_i\phi_i'(x) \cdot \sum_{j=1}^{n} u_j\phi_j'(x)dx. \tag{5.37}$$

Hence

$$\frac{\partial}{\partial u_i}J(u) = 0 \Rightarrow \sum_{j=1}^{n}\int_0^1 u_j\phi_j'(x)\phi_i'(x)dx = 0 \quad \forall i \in \{1,\ldots,n\}. \tag{5.38}$$

This gives exactly the same equations as in (5.22). Note the similarity with the result obtained by the use of Galerkin's method. Since Galerkin's method is applicable for more general cases, we do not treat this method further.

5.5 The treatment of a non-homogeneous Dirichlet boundary conditions

Suppose that we have to solve the following variational problem:

Find $u(x)$, subject to $u(0) = u_0$, such that

$$-\int_0^1 u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx \quad \forall v(0) = 0. \tag{5.39}$$

Now the essential condition is non-zero (when $u_0 \neq 0$), where u_0 is given. The treatment is similar to the case where $u_0 = 0$, but now we set

$$\bar{u}_n(x) = \sum_{j=0}^{n} u_j\phi_j(x) = u_0\phi_0(x) + \sum_{j=1}^{n} u_j\phi_j(x). \tag{5.40}$$

Note that u_0 is from the Dirichlet condition. Again, we use piecewise linear basis-functions $\{\phi_i(x)\}$, where

$$\phi_i(x_j) = \begin{cases} 1 & i = j, \\ 0, & i \neq j, \end{cases} \quad \text{and } \phi_i(x) \text{ is piecewise linear.} \tag{5.41}$$

For $\phi_0(x)$ we have: $\phi_0(x_0) = \phi_0(0) = 1$. For the functions $v(x)$ we set $\phi_i(x)$ where $i \in \{1,\ldots,n\}$ (note that $\phi_i(0) = 0$ since $v(0) = 0$). Then, we obtain the following problem:

Find $\{u_1, u_2, \ldots, u_n\}$ such that

$$-\int_0^1 \sum_{j=0}^{n} u_j\phi_j'(x)\phi_i'(x)dx = \int_0^1 f(x)\phi_i(x)dx, \tag{5.42}$$

for all $\{\phi_i(x)\}_{i=1}^n$, *i.e.* $\forall i \in \{1,\ldots,n\}$.

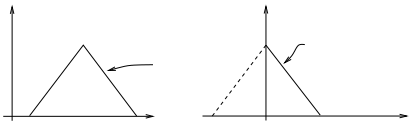


Figure 5.5: *The piecewise linear functions $\phi_i(x)$ and $\phi_0(x)$*

For the functions $\phi_i(x)$ we use the sketches from Figure 5.5. For $\phi_1(x)$, where $j = 1$, follows (where u_0 is prescribed):

$$-\int_0^1 u_0\phi_0'\phi_1'dx - \int_0^1 u_1\phi_1'\phi_1'dx - \int_0^1 u_2\phi_2'\phi_1'dx = \int_0^1 f(x)\phi_1 dx, \tag{5.43}$$

$$\left\{-\frac{1}{x_1 - x_0} - \frac{1}{x_2 - x_1}\right\} u_1 + \frac{1}{x_2 - x_1}u_2 = -\frac{u_0}{x_1 - x - 0} + f(x_1)\frac{1}{2}(x_2 - x_0). \tag{5.44}$$

Observe that u_0 appears in the right-hand side of the equation for $j = 1$. The integrals can also be evaluated for the other values of j. The same equations follow as in the example where $u_0 = 0$. Again, we will have that $\bar{u}_n'(1) \to 0$ as $n \to \infty$. For more dimensional problems, the same occurs: Suppose that we search u, subject to $u \mid_{\Gamma_2} = g$, such that

$$\int_\Omega \bar{\nabla}u \cdot \bar{\nabla}v dA = 0 \quad \forall v \mid_\Gamma = 0. \tag{5.45}$$

Let n be the number of gridnodes that are inside Ω (figure 5.5) or on boundary Γ_2, (but not on Γ_1), and let $\{n+1, \ldots, n+m\}$ be the gridnodes on boundary Γ_1, then we set

$$u(x,y) \approx \sum_{j=1}^{n} u_j\phi_j(x,y) + \sum_{j=n+1}^{n+m} g(x_j, y_j)\phi_j(x,y). \tag{5.46}$$

Here we take

$$\phi_j(x_i, y_i) = \begin{cases} 1, & \text{for } j = i, \\ 0, & \text{for } j \neq i, \end{cases} \tag{5.47}$$

where $\phi_j(x, y)$ are taken to be piecewise linear. Expression (5.46) is substituted into the weak form to obtain a system of linear equations.

5.6 A time-dependent example

Consider the following problem:

$$\begin{cases} \dfrac{\partial c}{\partial t} = \dfrac{\partial^2 c}{\partial x^2} \quad \text{on } x \in (0,1), t > 0, \\ c(0,t) = 1 \quad t > 0, \\ \dfrac{\partial c}{\partial x}(1,t) = 0 \quad t > 0, \\ c(x,0) = 0 \quad \text{on } x \in (0,1). \end{cases} \tag{5.48}$$

First we search a weak form for (5.48): Search c, $c(0,t) = 1$, $c(x,0) = 0$,

$$\int_0^1 \frac{\partial c}{\partial t} v dx = \int_0^1 \frac{\partial^2 c}{\partial x^2} v dx \quad \forall v(0,t) = 0. \tag{5.49}$$

To reduce the order of the derivative, we integrate by parts:

$$\int_0^1 \frac{\partial c}{\partial t} v dx = \left[\frac{\partial c}{\partial x} v\right]_0^1 - \int_0^1 \frac{\partial c}{\partial x}\frac{\partial v}{\partial x} dx = -\int_0^1 \frac{\partial c}{\partial x}\frac{\partial v}{\partial x} dx, \tag{5.50}$$

since $v(0,t) = 0$ and $\dfrac{\partial c}{\partial x}(1,t) = 0$. Then, we obtain the following weak form

$$\begin{cases} \text{Find } c, \text{ subject to } c(0,t) = 1,\ c(x,0) = 0, \text{ such that} \\ \displaystyle\int_0^1 \frac{\partial c}{\partial t} v dx = -\int_0^1 \frac{\partial c}{\partial x}\frac{\partial v}{\partial x} dx \quad \forall v(0) = 0. \end{cases} \tag{5.51}$$

We solve (5.51) by use of the Galerkin FEM. Again, we use piecewise linear basis-functions $\{\phi_i(x)\}_{i=0}^n$ as before on the n gridnodes, with

$$\phi_i(x_j) = \begin{cases} 1, & j = i, \\ 0, & j \neq i. \end{cases} \tag{5.52}$$

Then, we approximate $c(x,t)$ by

$$\bar{c}_n(x,t) = \sum_{j=0}^n c_j(t)\phi_j(x) = \phi_0(x) + \sum_{j=1}^n \phi_j(x) c_j(t). \tag{5.53}$$

Note that $\{c_j\}$ should be functions of x and t since $c(x,t)$ is a function of t and $\phi_j(x)$ is a function of x only. Substitution into (5.51) gives

$$\begin{cases} \text{Find } \{c_i(t)\}_{i=1}^n \text{ such that} \\ \int_0^1 \sum_{j=1}^n c_j'(t)\phi_j(x)\phi_i(x)dx \\ \quad = -\int_0^1 \left\{ \phi_0'(x) + \sum_{j=1}^n \phi_j'(x)c_j(t) \right\} \phi_i'(x)dx \quad \forall i \in \{1,\ldots,n\}. \end{cases} \tag{5.54}$$

The above problem represents a system of linear ordinary differential equations. Note that $c_j'(t) \approx \frac{c_j(t+\Delta t) - c_j(t}{\Delta t}$ We will deal with the solution of the time-dependent problem in the next chapter.

Exercise 5.6.1. *Extend the above derivation to a boundary value problem with time-dependent Robin boundary conditions imposed on the boundary $x = 1$ (cfr. Exercise 4.1.1)*

5.7 The principle of element matrices and vectors

We briefly treat the concept of element matrices, which is not a mathematical feature but only a convenient programming trick. This trick is used in most of the implementations of Finite Elements software.

For the treatment we treat a simple one-dimensional example:

$$\begin{cases} -u'' = f, \\ u(0) = u_0, \\ u'(1) = 0. \end{cases} \tag{5.55}$$

A weak form is then obtained by

$$\begin{aligned} & -u''v = fv \Leftrightarrow -\int_0^1 u''vdx = \int_0^1 fvdx \\ \Leftrightarrow \quad & \left[-u'v\right]_0^1 + \int_0^1 u'v'dx = \int_0^1 fvdx \\ \Leftrightarrow \quad & \int_0^1 u'v'dx = \int_0^1 fvdx. \end{aligned} \tag{5.56}$$

We use Galerkin's method on n gridpoints for the unknowns:

$$u(x) = \sum_{j=1}^n u_j\phi_j(x) + u_0\phi_0(x). \tag{5.57}$$

Substitution into the weak form, with

$$\phi_j(x_i) = \begin{cases} 1, & j = i, \\ 0, & j \neq i, \end{cases} \tag{5.58}$$

piecewise linear, gives

$$\int_0^1 \sum_{j=0}^{n} u_j \phi_j'(x)\phi_i'(x)dx = \int_0^1 f(x)\phi_i(x)dx, \quad i \in \{1, \ldots, n\}, \tag{5.59}$$

or

$$\sum_{j=1}^{n} u_j \int_0^1 \phi_j'(x)\phi_i'(x)dx = \int_0^1 f(x)\phi_i(x)dx - u_0 \int_0^1 \phi_0'(x)\phi_i'(x)dx \quad i \in \{1, \ldots, n\}. \tag{5.60}$$

In other words:

$$\begin{aligned} u_1 \int_0^1 \phi_1'\phi_1'dx + u_2 \int_0^1 \phi_2'\phi_1'dx + \cdots + u_n \int_0^1 \phi_n'\phi_1'dx &= \int_0^1 f\phi_1 - u_0 \int_0^1 \phi_0'\phi_1'dx, \\ u_1 \int_0^1 \phi_1'\phi_2'dx + u_2 \int_0^1 \phi_2'\phi_2'dx + \cdots + u_n \int_0^1 \phi_n'\phi_2'dx &= \int_0^1 f\phi_2 - u_0 \int_0^1 \phi_0'\phi_2'dx, \\ &\vdots \end{aligned}$$

For simplicity we take n=3; then,

$$\begin{aligned} u_1 \int_0^1 \phi_1'\phi_1'dx + u_2 \int_0^1 \phi_2'\phi_1'dx + u_3 \int_0^1 \phi_3'\phi_1'dx &= \int_0^1 f\phi_1 - u_0 \int_0^1 \phi_0'\phi_1'dx, \\ u_1 \int_0^1 \phi_1'\phi_2'dx + u_2 \int_0^1 \phi_2'\phi_2'dx + u_3 \int_0^1 \phi_3'\phi_2'dx &= \int_0^1 f\phi_2 - u_0 \int_0^1 \phi_0'\phi_2'dx, \\ u_1 \int_0^1 \phi_1'\phi_3'dx + u_2 \int_0^1 \phi_2'\phi_3'dx + u_3 \int_0^1 \phi_3'\phi_3'dx &= \int_0^1 f\phi_3 - u_0 \int_0^1 \phi_0'\phi_3'dx. \end{aligned} \tag{5.61}$$

We take piecewise linear basis-functions. Note that we have $\int_0^1 \phi_1'\phi_3'dx = 0$ (the basis-functions are called *nearly orthogonal*) for instance and that only contributions from neighbouring elements are non-zero. Next, we will introduce the concept element as an interval between adjacent meshpoint: $e_i := [x_{i-1}, x_i]$. Hence, using the fact that the basis-functions are nearly orthogonal, we write the

above equations as follows:

$$\begin{aligned} u_1 \int_{e_1 \cup e_2} \phi_1' \phi_1' dx + u_2 \int_{e_2} \phi_1' \phi_2' dx &= \int_{e_1 \cup e_2} f\phi_1 dx - u_0 \int_{e_1} \phi_0' \phi_1' dx, \\ u_1 \int_{e_2} \phi_2' \phi_1' dx + u_2 \int_{e_2 \cup e_3} \phi_2' \phi_2' dx + u_3 \int_{e_3} \phi_2' \phi_3' dx &= \int_{e_2 \cup e_3} f\phi_2 dx, \\ u_2 \int_{e_3} \phi_2' \phi_3' dx + u_3 \int_{e_3} \phi_3' \phi_3' dx &= \int_{e_3} f\phi_3 dx. \end{aligned} \tag{5.62}$$

Note further that

$$\begin{cases} \int_{e_i \cup e_{i+1}} \phi_i' \phi_i' dx = \int_{e_i} \phi_i' \phi_i' dx + \int_{e_{i+1}} \phi_i' \phi_i' dx \\ \int_{e_i \cup e_{i+1}} f\phi_i dx = \int_{e_i} f\phi_i dx + \int_{e_{i+1}} f\phi_i dx \end{cases} \tag{5.63}$$

Now we show a computer procedure to generate the stiff-ness matrix (discretization matrix) by use of element-matrices: start with,

$$A_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{5.64}$$

we finish with

$$A = \begin{pmatrix} \int_{e_1} \phi_1' \phi_1' + \int_{e_2} \phi_1' \phi_1' & \int_{e_2} \phi_1' \phi_2' & 0 \\ \int_{e_2} \phi_2' \phi_1' & \int_{e_2} \phi_2' \phi_2' + \int_{e_3} \phi_2' \phi_2' & \int_{e_3} \phi_2' \phi_3' \\ 0 & \int_{e_3} \phi_2' \phi_3' & \int_{e_3} \phi_3' \phi_3' \end{pmatrix}. \tag{5.65}$$

Now we introduce the concept of element-matrices:

$$s_{e_i} = \begin{bmatrix} \int_{e_i} \phi_{i-1}' \phi_{i-1}' dx & \int_{e_i} \phi_{i-1}' \phi_i' dx \\ \int_{e_i} \phi_i' \phi_{i-1}' dx & \int_{e_i} \phi_i' \phi_i' dx \end{bmatrix}. \tag{5.66}$$

Contributions for essential boundary conditions do not occur in the stiffness-matrix, hence for e_1 we obtain:

$$s_{e_1} = \left[\int_{e_1} \phi_1' \phi_1' dx \right], \tag{5.67}$$

and for e_2 and e_3:

$$s_{e_2} = \begin{bmatrix} \int_{e_2} \phi_1' \phi_1' dx & \int_{e_2} \phi_1' \phi_2' dx \\ \int_{e_2} \phi_2' \phi_1' dx & \int_{e_2} \phi_2' \phi_2' dx \end{bmatrix}, \tag{5.68}$$

$$s_{e_3} = \begin{bmatrix} \int_{e_3} \phi_2' \phi_2' dx & \int_{e_3} \phi_2' \phi_3' dx \\ \int_{e_3} \phi_3' \phi_2' dx & \int_{e_3} \phi_3' \phi_3' dx \end{bmatrix}. \tag{5.69}$$

The matrices s_{e_1}, s_{e_2} and s_{e_3} are the element-matrices of elements e_1, e_2 and e_3. We will put these matrices into A_0 and add them to obtain A. The position of the element-matrices is such that $s_{e_1}[1,1] = A_0[1,1]$, hence:

$$A_1 = \begin{pmatrix} \int\limits_{e_1} \phi_1'\phi_1' dx & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \tag{5.70}$$

Subsequently we add s_{e_2} to A_1 such that $s_{e_2}[2,2] = A_0[2,2]$:

$$\begin{aligned} A_2 &= \begin{pmatrix} \int\limits_{e_1} \phi_1'\phi_1' dx & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} \int\limits_{e_2} \phi_1'\phi_1' dx & \int\limits_{e_2} \phi_1'\phi_2' dx & 0 \\ \int\limits_{e_2} \phi_2'\phi_1' dx & \int\limits_{e_2} \phi_2'\phi_2' dx & 0 \\ 0 & 0 & 0 \end{pmatrix} \\ &= \begin{pmatrix} \int\limits_{e_1} \phi_1'\phi_1' dx + \int\limits_{e_2} \phi_1'\phi_1' dx & \int\limits_{e_2} \phi_1'\phi_2' dx & 0 \\ \int\limits_{e_2} \phi_2'\phi_1' dx & \int\limits_{e_2} \phi_2'\phi_2' dx & 0 \\ 0 & 0 & 0 \end{pmatrix}. \end{aligned} \tag{5.71}$$

Subsequently we add s_{e_3} to A_2 such that $s_{e_3}[2,2] = A_0[3,3]$ to obtain A_3:

$$\begin{aligned} A_3 &= \begin{pmatrix} \int\limits_{e_1} \phi_1'\phi_1' dx + \int\limits_{e_2} \phi_1'\phi_1' dx & \int\limits_{e_2} \phi_1'\phi_2' dx & 0 \\ \int\limits_{e_2} \phi_2'\phi_1' dx & \int\limits_{e_2} \phi_2'\phi_2' dx & 0 \\ 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \int\limits_{e_3} \phi_2'\phi_2' dx & \int\limits_{e_3} \phi_2'\phi_3' dx \\ 0 & \int\limits_{e_3} \phi_3'\phi_2' dx & \int\limits_{e_3} \phi_3'\phi_3' dx \end{pmatrix} \\ &= \begin{pmatrix} \int\limits_{e_1} \phi_1'\phi_1' dx + \int\limits_{e_2} \phi_1'\phi_1' dx & \int\limits_{e_2} \phi_1'\phi_2' dx & 0 \\ \int\limits_{e_2} \phi_2'\phi_1' dx & \int\limits_{e_2} \phi_2'\phi_2' dx + \int\limits_{e_3} \phi_2'\phi_2' dx & \int\limits_{e_3} \phi_2'\phi_3' dx \\ 0 & \int\limits_{e_3} \phi_3'\phi_2' dx & \int\limits_{e_3} \phi_3'\phi_3' dx \end{pmatrix}. \end{aligned} \tag{5.72}$$

Now, the stiffness-matrix has been built. The same principle is carried out with the element-vectors to generate the right-hand side vector. The same principle gives

$$\underline{f}_{e_i} = \begin{bmatrix} \int\limits_{e_i} f(x)\phi_{i-1}(x) dx \\ \int\limits_{e_i} f(x)\phi_i(x) dx \end{bmatrix}. \tag{5.73}$$

This is called the element-vector of element e_i.

$$\underline{f}_{e_1} = \left[\int\limits_{e_1} f(x)\phi_1(x) dx \right] \Rightarrow \underline{f}_1 = \begin{pmatrix} \int\limits_{e_1} f(x)\phi_1(x) dx \\ 0 \\ 0 \end{pmatrix}. \tag{5.74}$$

subsequently

$$\underline{f}_{e_2} = \begin{bmatrix} \int\limits_{e_2} f(x)\phi_1(x)dx \\ \int\limits_{e_2} f(x)\phi_2(x)dx \end{bmatrix} \tag{5.75}$$

$$\Rightarrow \underline{f}_2 = \underline{f}_1 + \begin{pmatrix} \int\limits_{e_2} f(x)\phi_1(x)dx \\ \int\limits_{e_2} f(x)\phi_2(x)dx \\ 0 \end{pmatrix} \tag{5.76}$$

$$= \begin{pmatrix} \int\limits_{e_1} f(x)\phi_1(x)dx + \int\limits_{e_2} f(x)\phi_1(x)dx \\ \int\limits_{e_2} f(x)\phi_2(x)dx \\ 0 \end{pmatrix}. \tag{5.77}$$

Subsequently

$$\underline{f}_{e_3} = \begin{bmatrix} \int\limits_{e_3} f(x)\phi_2(x)dx \\ \int\limits_{e_3} f(x)\phi_3(x)dx \end{bmatrix} \Rightarrow \underline{f}_3 = \begin{pmatrix} \int\limits_{e_1\cup e_2} f(x)\phi_1(x)dx \\ \int\limits_{e_2\cup e_3} f(x)\phi_2(x)dx \\ \int\limits_{e_3} f(x)\phi_3(x)dx \end{pmatrix}. \tag{5.78}$$

Now we have the right-handside if we take the essential condition into account for element e_1: $\begin{bmatrix} -u_0 \int\limits_{e_1} \phi_0'\phi_1' dx \end{bmatrix}$ at $\underline{f}_3'$, to obtain:

$$\underline{f} = \begin{pmatrix} \int\limits_{e_1\cup e_2} f(x)\phi_1(x)dx - u_0 \int\limits_{e_1} \phi_0'\phi_1' dx \\ \int\limits_{e_2\cup e_3} f(x)\phi_2(x)dx \\ \int\limits_{e_3} f(x)\phi_3(x)dx \end{pmatrix}. \tag{5.79}$$

The principle of element-matrices and vectors should be seen as a convenient programming trick. This trick allows constitutes the gateway for the implementation of the finite element method for problems with odd-shaped domains, local mesh refinement, unstructured grids, jumps in coefficients or problems requiring the use of parallel computing.

5.8 Numerical integration

Consider the following weak form

Find $u(x)$, subject to $u(0) = u_0$, such that

$$-\int_0^1 u'(x)v'(x)dx = \int_0^1 f(x)v(x)dx \quad \forall v(0) = 0. \tag{5.80}$$

Sometimes, the right-hand side contains a complicated integrand that cannot be integrated in terms of an elementary function. Since one uses the concept of element vectors and element matrices, one has to integrate the functions over an element. Many numerical integration methods are available, such as the midpoint rule, the trapezoidal rule, the Simpson rule and so on. A very common numerical integration rule based on the expression of a function in terms of a linear combination of basis-function is the *Newton-Cotes* integration Rule. In this section, we will consider Newton-Cotes integration with piecewise linear basis-functions. This rule is based on the following: Consider a function $g(x)$ to be integrated over an element $e_i := [x_{i-1}, x_i]$:

$$\int_{e_i} g(x)dx, \tag{5.81}$$

then we express $g(x)$ as a linear combination of basis-functions with the characteristics as mentioned before on the nodes x_{i-1} and x_i:

$$g(x) \approx g(x_{i-1})\phi_{i-1}(x) + g(x_i)\phi_i(x). \tag{5.82}$$

Then the integration over the interval zero-one, gives:

$$\int_{e_i} g(x)dx \approx g(x_{i-1}) \int_{e_i} \phi_{i-1}(x)dx + g(x_i) \int_{e_i} \phi_i(x)dx = \frac{g(x_{i-1}) + g(x_i)}{2}(x_i - x_{i-1}). \tag{5.83}$$

Note that the above integration formula represents the Trapezoidal Rule for numerical integration. The same is done with the right-hand of the above weak form:

$$\int_{e_i} f(x)\phi_k(x)dx \approx \sum_{j=i-1}^{i} f(x_j)\phi_k(x_j) \int_{e_i} \phi_j(x)dx = \frac{1}{2} f(x_k)(x_{k+1} - x_{k-1}), \tag{5.84}$$

for $k \in \{i-1, i\}$, note that $\phi_k(x_i) = 1$ if $i = k$ and else $\phi_k(x_i) = 0$. This integration rule is easily extended to more dimensional problems. This is beyond the scope of the present course. As an alternative, Gauss integration formulas can be used. This is not treated in this course.

5.9 Error considerations

This section is not necessary to understand the implementation of the finite element method. It is intended for the interested reader. The treatment of the error of the finite element solution will not be mathematical, but it will give an idea of what the error is and what issues are important for the derivation of the error. Suppose that we solve the equation in the following weak form of Poisson's equation

$$\begin{gathered} \text{Find } u \in H^1(\Omega), \text{ subject to } u = g \text{ on } \partial\Omega, \text{ such that} \\ \int_\Omega \nabla u \cdot \nabla \phi d\Omega = \int_\Omega \phi f d\Omega, \text{ for all } \phi \in H^1(\Omega). \end{gathered} \tag{5.85}$$

To solve this problem using Galerkin's method, we approximate the solution by $u(\mathbf{x}) \approx u_h(\mathbf{x}) = \sum_{j=1}^{n} u_j \phi_j(\mathbf{x})$, where $\phi_j(\mathbf{x}) = 0$ on $\partial\Omega$. Here $u(\mathbf{x})$ represents the exact solution and u_j represents

the approximate solution at the nodal points. For the approximate solution u_h, we have

$$\begin{aligned} &\text{Find } u_h \in H_h^1(\Omega), \text{ subject to } u_h = g \text{ on } \partial\Omega, \text{ such that} \\ &\int_\Omega \nabla u_h \cdot \nabla \phi_h d\Omega = \int_\Omega \phi_h f d\Omega, \text{ for all } \phi_h \in H_h^1(\Omega), \end{aligned} \tag{5.86}$$

where $H_h^1(\Omega)$ represents the set of solutions for the approximate solution u_h. Note that $H_h^1(\Omega) \subset H^1(\Omega)$. Let $\tilde{u}(\mathbf{x})$ be the approximate solution with the exact values of the solution $u(\mathbf{x})$, that is

$$\tilde{u}(\mathbf{x}) = \sum_{j=1}^{n} u(\mathbf{x}_j)\phi_j(\mathbf{x}).$$

Note that the difference between the exact solution $u(\mathbf{x})$ and the above solution $\tilde{u}(\mathbf{x})$ is only determined by the interpolation method that is used. Then it can be proved (see Braess [1996] for instance) that

$$\int_\Omega ||\nabla(u - u_h)||^2 d\Omega \leq \int_\Omega ||\nabla(u - \tilde{u})||^2 d\Omega. \tag{5.87}$$

The above integrals represent errors in the energy-norm. The left-hand side of the above inequality gives the total error of the finite element solution with respect to the exact solution in the so-called energy norm. This total error basically has two sources:

[1] A finite set of basis functions (based on the finite number of meshpoints) is chosen, and hence the summation only concerns a finite number of terms;

[2] Using interpolation functions for the basis functions, this gives an interpolation error, which depends on the order of the interpolation functions $\phi_j(\mathbf{x})$.

The right-hand side of the above inequality only concerns the interpolation error. The inequality (5.87) is very convenient since it says in the energy norm that the total error is bounded from above by the interpolation error. This last mentioned error depends on the polynomial order. If linear elements are used, then it can be demonstrated that the energy norm of the interpolational error is of order $O(h)$ where h represents the largest side of the element. The actual error $u - u_h$ is one order higher, that is $O(h^2)$. The reader should realize that many statements in this section have been made while omitting some subtle mathematical issues.

As a rule of thumb, we use that if the interpolational error is of order $O(h^p)$, then the order of the error, that is the difference between the finite element solution and the exact solution is of order $O(h^{p+1})$. It should be mentioned that this fact only holds if the elements are not degenerate (the vertices of the triangle are not located on a line in the case of triangular elements).

6

Time Dependent Problems: Numerical Methods

In many cases the mathematical models are *time-dependent*, by which we mean that their solution depends on time. Typical examples are

$$\frac{\partial c}{\partial t} + \bar{q} \cdot \bar{\nabla} c = \Delta c \qquad \text{(convection-diffusion)} \tag{6.1}$$

$$\frac{\partial c}{\partial t} - \Delta c = 0 \qquad \text{(transient diffusion)} \tag{6.2}$$

$$\frac{\partial^2 c}{\partial t^2} = \Delta c \qquad \text{(wave transmission)} \tag{6.3}$$

$$\frac{\partial c}{\partial t} + \frac{\partial}{\partial x} f(c) = 0 \qquad \text{(Buckley-Leverett in 2-plane flow)} \tag{6.4}$$

In this chapter we will treat some time-integration methods in relation to Finite Diffrences and Finite Elements. Important subjects will be consistency (convergence), accuracy and stability.

6.1 Time-integration methods

First we consider an ODE (Ordinary Differential Equation) to illustrate some time-integration methods. Subsequently, we will apply these methods to PDE's. Consider the following problem:

$$\begin{cases} y(0) = y_0 \quad t = 0, \\ y'(t) = f(y,t) \quad t > 0. \end{cases} \tag{6.5}$$

Let T_{end} be the end-time of the numerical simulation $\Delta t = \dfrac{T_{\text{end}}}{m}$, m= number of timesteps, then we introduce the notation $y^n = y(t_n) = y(n\Delta t)$ and u^n for the exact solution evaluated in the discrete time steps and its numerical approximation, respectively. At $t = 0$ the relation $u^0 = y^0$ holds. We will formulate some classical time-integration methods:

[1] Euler's forward time integration method (explicit):

$$u^{n+1} = u^n + \Delta t f(u^n, t_n) \quad n \geq 0 \quad n \in \{0, \ldots, m-1\} \tag{6.6}$$

[2] Heun's (or improved Euler's) time integration method

$$\begin{cases} \bar{u}^{n+1} = u^n + \Delta t f(u^n, t_n) & \text{(predictor)} \\ u^{n+1} = u^n + \frac{\Delta t}{2}\left[f(u^n, t_n) + f(\bar{u}^{n+1}, t_{n+1})\right] & \text{(corrector)} \end{cases} \tag{6.7}$$

[3] Euler's backward time integration method (implicit)

$$u^{n+1} = u^n + \Delta t f(u^{n+1}, t^{n+1}) \tag{6.8}$$

(Note that we have to solve a non-linear equation whenever f is non-linear in u)

[4] Crank-Nicholson's time integration method

$$u^{n+1} = u^n + \frac{\Delta t}{2}\left[f(u^n, t_n) + f(u^{n+1}, t_{n+1})\right] \tag{6.9}$$

(Here again a non-linear problem has to be solved when f is nonlinear in u)

More methods for time-integration can be found in e.g. Burden and Faires [2001]; Vuik et al. [2007]. The four methods listed are the most common and we will analyse them. We also note that the modified Euler method falls within the class of the so-called multi-step methods (due to Runge-Kutta). These methods can be adjusted such that always the desired accuracy can be obtained.

6.2 Accuracy of time-integration methods

We first analyse the *local truncation* error, i.e., the error made in a single time step, of the Euler's forward method. Given, again

$$\begin{cases} y(0) = y_0 \quad t = 0, \\ y'(t) = f(y, t) \quad t > 0, \end{cases} \tag{6.10}$$

then by Taylor series expansion

$$\begin{aligned} y^{n+1} &= y(t_{n+1}) \\ &= y(t_n + \Delta t) \\ &= y(t_n) + \Delta t\, y'(t_n) + \frac{\Delta t^2}{2} y''(t_n) + \frac{\Delta t^3}{6} y'''(t_n) + \mathcal{O}(\Delta t^4). \end{aligned} \tag{6.11}$$

To analyse the local truncation, we assume the computed solution u^n at time step n has no error, i.e., $u^n = y^n$. This implies that for the next time step

$$\begin{aligned} u^{n+1} &= u^n + \Delta t\, f(u^n, t_n) \quad \text{(Euler forward)} \\ &= y^n + \Delta t\, f(y^n, t_n) \quad (u^n = y^n) \\ &= y^n + \Delta t\, y'(t_n)\,. \qquad (y'(t) = f(y, t) \quad t > 0) \end{aligned} \tag{6.12}$$

Subtracting (6.12) from equation (6.11) then yields

$$y^{n+1} - u^{n+1} = \frac{\Delta t^2}{2} y''(t_n) = \mathcal{O}(\Delta t^2). \tag{6.13}$$

Herewith, the local truncation error at step $n+1$ is given by:

$$\begin{aligned} \tau_{n+1}(\Delta t) &= \frac{y^{n+1} - u^{n+1}}{\Delta t} && (6.14)\\ &= \frac{\Delta t}{2} y''(t_n) = \mathcal{O}(\Delta t) && (6.15) \end{aligned}$$

where the factor $1/\Delta t$ in (6.14) is introduced to compensate for the fact that the scheme (6.6) integrated the ODE times Δt. Hence the local truncation error for the Euler's forward method is order of Δt. Similar analysis shows that the local truncation error for the Euler's backward method is also order of Δt.

Next we analyse the modified Euler method:

$$\left\{ \begin{array}{lcl} \bar{u}^{n+1} &=& u^n + \Delta t f(u^n, t_n), \\ u^{n+1} &=& u^n + \frac{\Delta t}{2}\left[f(u^n, t_n) + f(\bar{u}^{n+1}, t_{n+1})\right] \\ &=& u^n + \frac{\Delta t}{2}\left[f(u^n, t_n)\right] + \frac{\Delta t}{2}\left[f(u^n + \Delta t\, f(u^n, t_n), t_n + \Delta t)\right] \end{array} \right. \tag{6.16}$$

By expanding f in a first order Taylor sequence in both arguments, the last term in the right-hand side of the above expression can be written as

$$\begin{aligned} f(u^n + \Delta t\, f(u^n, t_n), t_n + \Delta t) &= f(u^n, t_n) + \Delta t\, f(u^n, t_n)\frac{\partial f}{\partial u}(u^n, t_n) \\ &\quad + \Delta t\, \frac{\partial f}{\partial t}(u^n, t_n) + \mathcal{O}(\Delta t^2) \end{aligned} \tag{6.17}$$

After collecting terms in $f(u^n, t_n)$, u^{n+1} can then be written as

$$\begin{aligned} u^{n+1} &= u^n + \Delta t f(u^n, t_n) + \frac{\Delta t^2}{2}\frac{\partial f}{\partial u}(u^n, t_n) f(u^n, t_n) + \frac{\Delta t^2}{2}\frac{\partial f}{\partial t}(u^n, t_n) \\ &\quad + \mathcal{O}(\Delta t^3) \end{aligned} \tag{6.18}$$

As before, we investigate the error made in a single time step and assume that $u^n = y^n$. This implies that (6.18) reduces

$$\begin{aligned} u^{n+1} &= y^n + \Delta t f(y^n, t_n) + \frac{\Delta t^2}{2}\frac{\partial f}{\partial y}(y^n, t_n) f(y^n, t_n) + \frac{\Delta t^2}{2}\frac{\partial f}{\partial t}(y^n, t_n) \\ &\quad + \mathcal{O}(\Delta t^3) \end{aligned} \tag{6.19}$$

from which f can be eliminated using

$$\begin{aligned} f(y^n, t_n) &= y'(t_n) \\ \frac{\partial f}{\partial y}(y^n, t_n) f(y^n, t_n) + \frac{\partial f}{\partial t}(y^n, t_n) &= \frac{\partial f}{\partial y}(y^n, t_n)\, y'(t_n) + \frac{\partial f}{\partial t}(y^n, t_n) \\ &= \frac{df}{dt}(y^n, t_n) \quad \text{(chain rule)} \\ &= y''(t_n) \end{aligned} \tag{6.20}$$

resulting in

$$u^{n+1} = y^n + \Delta t\, y'(t^n) + \frac{\Delta t^2}{2} y''(t_n) + \mathcal{O}(\Delta t^3)\,. \tag{6.21}$$

Herewith, the local truncation error is given by:

$$\tau_{n+1}(\Delta t) = \frac{y^{n+1} - u^{n+1}}{\Delta t} = \mathcal{O}(\Delta t^2). \tag{6.22}$$

Hence the Modified Euler method has a local truncation error of order Δt^2.

Exercise 6.2.1. *Use an analysis as above to show that the local truncation error of the Crank-Nicholson scheme is* $\mathcal{O}(\Delta t^2)$.

Next we analyse the *global* error, i.e, the error accumulated by integrating from $t = 0$ to $t = T_{end}$. This error is computed by summing up the contributions $y^n - u^n$ at each time step. If the interval $[0, T_{end}]$ is divided into m steps, then $m = T_{end}/\Delta t = \mathcal{O}(\Delta t^{-1})$. If therefore the local contribution $y^n - u^n$ is $\mathcal{O}(\Delta t^2)$ (Euler's forward or backward) and $\mathcal{O}(\Delta t^3)$ (Modified Euler or Crank-Nicolson), then the global error is $\mathcal{O}(\Delta t)$ (Euler's forward or backward) and $\mathcal{O}(\Delta t^2)$ (Modified Euler or Crank-Nicholson).

6.3 Time-integration of PDE's

The time integration method for PDE's is analogous to ODE's. Suppose that the following problem is given:

$$\begin{cases} \dfrac{\partial u}{\partial t} = \mathcal{L}(u), \\ u(x,0) = u_0, \\ \text{with appropriate (BC)}, \end{cases} \quad , \tag{6.23}$$

where $\mathcal{L}(u)$ represents a linear space differential operator; example:

$$\mathcal{L}(u) = \Delta u \quad \text{(diffusion), or,} \tag{6.24}$$

$$\mathcal{L}(u) = \Delta u - \bar{q} \cdot \bar{\nabla} u \quad \text{(convection-diffion).} \tag{6.25}$$

$$\text{and so on...} \tag{6.26}$$

The time discretization part $\dfrac{\partial u}{\partial t}$ is done (for instance) by:

$$\frac{\partial u}{\partial t} = \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\Delta t}, \tag{6.27}$$

where $u_{i,j}^n$ denotes $u(x_i, y_j, t^n)$. Then we give the following example:

$$\begin{cases} u_{i,j}^{n+1} = u_{i,j}^n + \Delta t \mathcal{L}(u^n) \quad \text{Forward(Explicit) Euler,} \\ u_{i,j}^{n+1} = u_{i,j}^n + \Delta t \mathcal{L}(u^{n+1}) \quad \text{Backward (Implicit) Euler,} \\ u_{i,j}^{n+1} = u_{i,j}^n + \frac{\Delta t}{2}\left\{\mathcal{L}(u^{n+1}) + \mathcal{L}(u^n)\right\} \quad \text{Crank-Nicholson,} \\ \begin{cases} \bar{u}_{i,j}^{n+1} = u_{i,j}^n + \Delta t \mathcal{L}(u^n) \quad \text{Modified Euler (predictor-corrector),} \\ u_{i,j}^{n+1} = u_{i,j}^n + \frac{\Delta t}{2}\left\{\mathcal{L}(u^n) + \mathcal{L}(\bar{u}^{n+1})\right\}. \end{cases} \end{cases} \tag{6.28}$$

The global errors are $\mathcal{O}(\Delta t)$, $\mathcal{O}(\Delta t)$, $\mathcal{O}(\Delta t^2)$ and $\mathcal{O}(\Delta t^2)$, respectively. This can be shown by the use of similar procedures as in Section 6.2. Now we consider the application to a diffusion equation as our example. We will illustrate the use of finite differences and finite flements for a one-dimensional diffusion problem and a 1-dimensional wave equation.

6.3.1 The heat equation

We consider the discretization of the heat equation

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < 1, \quad t > 0 \tag{6.29}$$

$$+ \text{ Dirichlet boundary conditions } + \text{ initial condition.}$$

[1] Finite Differences

$$\frac{\partial u}{\partial t} = \frac{u_{i+1} - 2u_i + u_{i-1}}{\Delta x^2}, \tag{6.30}$$

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \frac{u_{i+1}^n - 2u_i^n + u_{i-1}^n}{\Delta x^2} \quad \text{(Forward Euler)}, \tag{6.31}$$

$$\frac{u_i^{n+1} - u_i^n}{\Delta t} = \frac{u_{i+1}^{n+1} - 2u_i^{n+1} + u_{i-1}^{n+1}}{\Delta x^2} \quad \text{(Backward Euler)}. \tag{6.32}$$

[2] Finite Elements for (6.29) Weak form:

$$\int_\Omega \frac{\partial u}{\partial t} v dx = \int_\Omega \frac{\partial^2 u}{\partial x^2} v dx = -\int_\Omega \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} dx, \tag{6.33}$$

where we took $v\,|_{\partial\Omega} = 0$. Find u, subject to $u(0) = u_0$, $u(1) = u_1$, such that

$$\int_0^1 \frac{\partial u}{\partial t} v dx = -\int_0^1 \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} dx \quad \forall v(0) = 0 = v(1)\,, \tag{6.34}$$

provided that the above integral exist. Further take basis-functions (piecewise linear), $\phi_i(0) = 0 = \phi_i(1)$ for $i \in \{1, \ldots, m-1\}$,

$$u(x,t) = \sum_{j=1}^{m-1} [c_j(t)\phi_j(x)] + u_0\phi_0(x) + u_1\phi_m(x). \tag{6.35}$$

Take for simplicity $u_0 = 0 = u_1$ then

$$\sum_{j=1}^{m-1} c_j'(t) \int_0^1 \phi_j(x)\phi_i(x) dx = -\sum_{j=1}^{m-1} c_j(t) \int_0^1 \phi_j'(x)\phi_i'(x) dx \quad i \in \{1, \ldots, m-1\}. \tag{6.36}$$

For forward Euler, one obtains:

$$\sum_{j=1}^{m-1} \frac{c_j^{n+1} - c_j^n}{\Delta t} \int_0^1 \phi_j(x)\phi_i(x) dx = -\sum_{j=1}^{m-1} c_j^n \int_0^1 \phi_j'(x)\phi_i'(x) dx \quad i \in \{1, \ldots, m-1\}, \tag{6.37}$$

where we call $M_{ij} = \int_0^1 \phi_i(x)\phi_j(x)dx$ the entries of the mass-matrix and $S_{ij} = \int_0^1 \phi_i'(x)\phi_j'(x)dx$ the entries of the stiffnes-matrix. Elementwise, we obtain for $i \in \{1, .., m-1\}$

$$\begin{aligned} \frac{c_{i-1}^{n+1} - c_{i-1}^n}{\Delta t} \int_{e_i} \phi_i\phi_{i-1}dx \quad + \quad \frac{c_i^{n+1} - c_i^n}{\Delta t} \int_{e_i \cup e_{i+1}} (\phi_i)^2 dx + \frac{c_{i+1}^{n+1} - c_{i+1}^n}{\Delta t} \int_{e_{i+1}} \phi_i\phi_{i+1}dx \\ = -c_{i-1}^n \int_{e_i} \phi_i'\phi_{i-1}'dx \quad - \quad c_i^n \int_{e_i \cup e_{i+1}} \phi_i'\phi_i'dx - c_{i+1}^n \int_{e_{i+1}} \phi_i'\phi_{i+1}'dx\,. \end{aligned} \tag{6.38}$$

All the integrals can be computed. The discretization method for FEM looks very different. For the case of an equidistant grid, it can be shown that the result becomes identical to the finite difference (and finite volume) method. For backward Euler, one obtains similarly:

$$\begin{aligned} \frac{c_{i-1}^{n+1} - c_{i-1}^n}{\Delta t} \int_{e_i} \phi_i\phi_{i-1}dx \quad + \quad \frac{c_i^{n+1} - c_i^n}{\Delta t} \int_{e_i \cup e_{i+1}} (\phi_i)^2 dx + \frac{c_{i+1}^{n+1} - c_{i+1}^n}{\Delta t} \int_{e_{i+1}} \phi_i\phi_{i+1}dx \\ = -c_{i-1}^{n+1} \int_{e_i} \phi_i'\phi_{i-1}'dx \quad - \quad c_i^{n+1} \int_{e_i \cup e_{i+1}} \phi_i'\phi_i'dx - c_{i+1}^{n+1} \int_{e_{i+1}} \phi_i'\phi_{i+1}'dx \\ \text{for } i \in \{1, .., m-1\}. \end{aligned} \tag{6.39}$$

All these expressions can be adjusted to Crank-Nicholson or a two-step method (Runge-Kutta 2 or Modified Euler). Later we will consider the stability of numerical methods.

We remark further that the above finite element discretization of the heat problem can be written in matrix-form as

$$M\frac{d\underline{c}}{dt} = S\underline{c}, \tag{6.40}$$

where $M_{ij} = \int_0^1 \phi_i(x)\phi_j(x)dx$, $S_{ij} = \int_0^1 \phi_i'(x)\phi_j'(x)dx$ and $\underline{c} = [c_1 \dots c_n]^T$. Of course, the Euler, modified Euler *etc.* can be written similarly as before. As an example the Euler-forward time integration method can be written by:

$$M\underline{c}^{n+1} = M\underline{c}^n + \Delta t S\underline{c}^n. \tag{6.41}$$

For the case that M is not a diagonal matrix, a linear system of equations has to be solved also for the Euler-forward method.

Exercise 6.3.1. *Write down the corresponding time-integration methods for the backward Euler, Modified Euler and Crank-Nicholson methods.*

6.3.2 The wave equation

We consider the discretization of the wave equation

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2}, \quad 0 < x < 1, \quad t > 0,$$
$$+ \text{ Dirichlet boundary conditions } + \text{ initial conditions for } u \text{ and } u_t. \tag{6.42}$$

Here c represents the given wave-speed, which is assumed to be constant in the text. The weak form of the above equation is obtained after multiplication by a test-function, to obtain

$$\int_\Omega \frac{\partial^2 u}{\partial t^2} v dx = c^2 \int_\Omega \frac{\partial^2 u}{\partial x^2} v dx = -c^2 \int_\Omega \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} dx, \tag{6.43}$$

where we took $v \mid_{\partial\Omega} = 0$. Find u, subject to $u(0) = u_0$, $u(1) = u_1$, such that

$$\int_0^1 \frac{\partial^2 u}{\partial t^2} v dx = -c^2 \int_0^1 \frac{\partial u}{\partial x}\frac{\partial v}{\partial x} dx \quad \forall v(0) = 0 = v(1)\,, \tag{6.44}$$

provided that the above integrals exist. Further take basis-functions (piecewise linear), $\phi_i(0) = 0 = \phi_i(1)$ for $i \in \{1, \ldots, m-1\}$,

$$u(x,t) = \sum_{j=1}^{m-1} [c_j(t)\phi_j(x)] + u_0\phi_0(x) + u_1\phi_m(x). \tag{6.45}$$

Take for simplicity $u_0 = 0 = u_1$ then

$$\sum_{j=1}^{m-1} c_j''(t) \int_0^1 \phi_j(x)\phi_i(x) dx = -c^2 \sum_{j=1}^{m-1} c_j(t) \int_0^1 \phi_j'(x)\phi_i'(x) dx \quad i \in \{1, \ldots, m-1\}. \tag{6.46}$$

where we call $M_{ij} = \int_0^1 \phi_i(x)\phi_j(x) dx$ the entries of the mass-matrix and $S_{ij} = \int_0^1 \phi_i'(x)\phi_j'(x) dx$ the entries of the stiffnes-matrix. Hence, we obtain

$$M\frac{d^2\underline{c}}{dt^2} = c^2 S\underline{c}, \tag{6.47}$$

One possibility to integrate the above equation in time is to write it as system of first-order differential equations in time for c and w:

$$\frac{d\underline{c}}{dt} = \underline{w},$$
$$M\frac{d\underline{w}}{dt^2} = c^2 S\underline{c}, \tag{6.48}$$
$$\text{with initial conditions for } c \text{ and } w.$$

Now, we can apply the Forward Euler method to the above problem:

$$\begin{aligned} \underline{c}^{n+1} &= \underline{c}^n + \Delta t w^n; \\ M\underline{w}^{n+1} &= M\underline{w}^n + \Delta t c^2 S\underline{c}^n. \end{aligned} \tag{6.49}$$

Exercise 6.3.2. *Write down the corresponding time-integration methods for the backward Euler, Modified Euler and Crank-Nicholson methods.*

Exercise 6.3.3. *In case of an uniform grid, compute all integrals and show that the FEM and finite difference discretization do coincide.*

It can be shown that the numerical integration of the above equation gives rise to dissipation ($\underline{c}^n \to 0$ as $n \to \infty$). The amount of dissipation can be decreased when a Runge-Kutta method is used. Further one can use a direct time-integration of the second-order system of differential equations. This is not treated any further.

6.4 Stability analysis

In general after discretization we obtain a system of ordinary differential equation in the form

$$M\frac{d\underline{u}}{dt} = S\underline{u} \Leftrightarrow \frac{d\underline{u}}{dt} = M^{-1}S\underline{u}. \tag{6.50}$$

For the stability the eigenvalues of the above matrix $M^{-1}S$ are crucial. First, for the analytical asymptotic stability we require

$$\lambda(M^{-1}S) < 0, \text{ and if } \lambda \notin \mathbb{R} \text{ we have } \Re\{\lambda(M^{-1}S)\} < 0. \tag{6.51}$$

Exercise 6.4.1. *Derive stability criteria for the eigenvalues of $M^{-1}S$ for the Forward Euler, Modified Euler, Backward Euler and Trapezoidal time integration methods.*

Hence, it is crucially important to have some knowledge on the eigenvalues of the matrix $M^{-1}S$. For large matrices it is not easy to compute the eigenvalues. Fortunately, Gershgorin's Theorem gives a often very usefull estimate of the eigenvalues. For many cases the mass matrix M is diagonal as a result of numerical integration by use of the Rule of Newton-Cotes. This is commonly called *lumping*.

Exercise 6.4.2. *Show by the numerical integration of Newton-Cotes that*

$$m_{ji} = \int_0^1 \phi_i(x)\phi_j(x)dx = \begin{cases} \dfrac{x_{i+1} - x_{i-1}}{2}, & \text{if } j = i, \\ 0, & \text{else.} \end{cases} \tag{6.52}$$

For this case the Theorem reads as follows:

Theorem 6.4.1. *Let M be diagonal, then, for all eigenvalues λ of $M^{-1}S$ holds:*

$$|\lambda| \leq \sup_k \frac{1}{|m_{kk}|} \sum_{i=1}^{n} |s_{ki}|. \tag{6.53}$$

Note that the eigenvalues may be complex if S is not symmetric. Then the areas in which the eigenvalues are allowed to be in, consist of circles in the complex plane.

Proof 6.4.1. *Let λ be an eigenvalue of the generalized eigenvalue problem with correspoding eigenvector $\underline{v}$, then,*

$$S\underline{v} = \lambda M\underline{v}. \tag{6.54}$$

This immplies that for each component k:

$$\sum_{i=1}^{n} s_{ki}v_i = \lambda m_{kk}v_k. \tag{6.55}$$

Let v_j be the component of $\underline{v}$ with the largest modulus, then, we have for this index j:

$$\lambda = \frac{1}{m_{jj}} \sum_{i=1}^{n} s_{ji}\frac{v_i}{v_j}, \tag{6.56}$$

and since $|v_i/v_j| \leq 1$, we get

$$|\lambda| \leq \frac{1}{|m_{kk}|} \sum_{i=1}^{n} |s_{ki}|. \tag{6.57}$$

This proves the assertion. □

We illustrate the use of Gershgorin's Theorem by the following example.

Example: Suppose we use Finite Differences, then $M = I$. Let h be the stepsize and Δt be the time-step. Further in one dimension, we have $s_{ii} = -2/h^2$ and $s_{ii-1} = 1/h^2 = s_{ii+1}$. From this, we obtain $\lambda \leq 4/h^2$. Hence we obtain for the Forward Euler method: $\Delta t \leq h^2/2$.

The analysis of stability of time-integration methods for Partial Differential Equations can also be done by the use of the Von Neumann analysis, which is based on Fourier analysis. This is omitted in the present course.

Bibliography

M. Ainsworth and J.T. Oden. *A posteriori error estimation in finite element analysis*. Pure and applied mathematics. John Wiley, 2000. ISBN 9780471294115.

U.M. Ascher and L.R. Petzold. *Computer methods for ordinary differential equations and differential-algebraic equations*. Miscellaneous Titles in Applied Mathematics Series. Society for Industrial and Applied Mathematics, 1998. ISBN 9780898714128.

K. E. Atkinson and W. Han. *Theoretical Numerical Analysis: A Functional Analysis Framwork*. Number 39 in Texts in Applied Mathematics. Springer, second edition, 2005.

D. Braess. *Finite Elements*. Cambridge University Press, 1996.

S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Method*, volume 15 of *Texts in Applied Mathematics*. Springer, third edition, 2008.

R.L. Burden and J.D. Faires. *Numerical Analysis*. Brooks-Cole publishing company, seventh edition, 2001.

C. Cuvelier, A. Segal, and A.A. van Steenhoven. *Finite Element Methods and Navier-Stokes Equations*. D. Reidel Publising Company, Dordrecht, 1986.

J. E. Dennis Jr. and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, volume 16 of *Classics in Applied Mathematics*. SIAM, Philadelphia, PA, 1996.

K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. *Computational Differential Equations*. The Press Syndicate of the University of Cambridge, 1996.

L.C. Evans. *Partial Differential Equations*. AMS, third edition edition, 1999.

M. Farrashkhalvat and J. P Miles. *Basic Structured Grid Generation with Introduction to Unstructured Grid Generation*. Butterworth-Heinemann, 2003.

P. J. Frey and P. L. George. *Mesh Generation*. Hermes Science Publishing, Oxford, 2000.

G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, third edition, 1996.

C. T. Kelley. *Iteratieve Methods for Linear and Nonlinear Equations*, volume 16 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, PA, 1995. ISBN 0-89871-352-8.

E. Kreyszig. *Introductory functional analysis with applications*. Wiley, New York, 1989.

V. D. Liseikin. *Grid Generation Methods*. Springer-Verlag, Berlin, 2010.

K. W. Morton and D. F. Mayers. *Numerical Solution of Partial Differential Equations*. Cambridge University Press, 1994. ISBN 0-521-42922-6.

A. Quarteroni and A. Valli. *Numerical Approximation of Partial Differential Equations*. Springer-Verlag, Berlin, 1994.

Y. Saad. *Iterative Methods for Sparse Linear Systems*. PWS Publishing Company, Boston, MA, 1996. ISBN 0-534-94776-X.

B. F. Smith, P. O. Bjørstad, and W. D. Gropp. *Domain Decompostion: Parallel Multilevel Methods for Elliptic Partial Differential Equations*. Cambridge University Press, 1996.

G. Strang and G.J. Fix. *An analysis of the Finite Element Method*. Prentice-Hall Series, Englewood Cliffs, N.J, 1973.

U. Trottenberg, C. Oosterlee, and A. Schüller. *Multigrid*. Academic Press, San Diego, 2001.

H. A. van der Vorst. *Iterative Krylov Methods for Large Linear Systems*, volume 13 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, 2003.

J. van Kan, A. Segal, and F.J. Vermolen. *Numerical methods in scientific computing*. Delftse Uitgevers Maatschappij, 2006.

C. Vuik, P. van Beek, F. Vermolen, and J. van Kan. *Numerical Methods for Ordinary Differential Equations*. Delftse Uitgevers Maatschappij, 2007.

O. C. Zienkiewics and R. L. Taylor. *The Finite Element Method: Volume 1 The Basics*. Butterworth-Heinemann, fifth edition edition, 2000a.

O. C. Zienkiewics and R. L. Taylor. *The Finite Element Method: Volume 2 Solid Mechanics*. Butterworth-Heinemann, fifth edition edition, 2000b.