LTCC Module Computational Methods

Lecture Notes

Norbert Heuer, Brunel University

1 Introduction to the finite element method

1.1 Weak formulation of the Poisson equation (homogeneous Dirichlet)

Our model problem is the *Poisson equation* with homogeneous *Dirichlet boundary condition* on a sufficiently smooth, simply connected bounded domain $\Omega \subset \mathbf{R}^2$ (all stated results will be valid for Lipschitz domains):

$$\begin{aligned} -\Delta u &= f \qquad && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma. \end{aligned} \tag{1.1}$$

Here, f is a given function and Δ is the *Laplace operator* or *Laplacian* defined by

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}$$

where $x = (x_1, x_2) \in \mathbf{R}^2$ are Cartesian coordinates. Even though the Poisson equation looks very special it is an important model case representing several problems from physics based on energy minimisation. Variations of the techniques we will study apply to a wide class of second order so-called *elliptic* problems.

It is known that there are cases where no classical (i.e. twice continuously differentiable) solution of (1.1) exists. In order to deal with a uniquely solvable problem one therefore derives a weaker formulation.

It is convenient to write the Laplace operator in the following form:

$$\Delta u = \operatorname{div} \nabla u$$

where ∇u is the *gradient* of u defined by

$$\nabla u = \left(\frac{\partial u}{\partial x_1}, \frac{\partial u}{\partial x_2}\right) \quad \text{in } \Omega$$

and div is the *divergence* defined for a vector-valued function $A = (A_1, A_2)$ by

$$\operatorname{div} A = \frac{\partial A_1}{\partial x_1} + \frac{\partial A_2}{\partial x_2} \quad \text{in } \Omega.$$

We will also need the *normal derivative* of a function w defined by

$$\partial_n w := \frac{\partial w}{\partial n} := n \cdot \nabla w = \frac{\partial w}{\partial x_1} n_1 + \frac{\partial w}{\partial x_2} n_2 \quad \text{on } \Gamma.$$

Here, $n = (n_1, n_2)$ denotes the exterior unit normal vector along Γ.

Recall the following integration-by-parts formula.

Lemma 1.1 (First Green formula) *For sufficiently smooth functions v and $w = (w_1, w_2)$ there holds*

$$\int_\Omega \nabla v \cdot w \, dx = \int_\Gamma v \, n \cdot w \, ds - \int_\Omega v \operatorname{div} w \, dx. \tag{1.2}$$

The first integral on the right-hand side denotes integration with respect to the arc length s along Γ.

Remark 1.1 *Remember that, for a differentiable curve Γ with parameter representation $\gamma = (\gamma_1, \gamma_2) : (0, R) \to \Gamma \subset \mathbf{R}^2$, integration along Γ with respect to the arc length is defined by*

$$\int_\Gamma f \, ds = \int_0^R f(\gamma(t)) \left| \frac{d\gamma}{dt}(t) \right| \, dt = \int_0^R f(\gamma(t)) \sqrt{\Big(\frac{d\gamma_1(t)}{dt}\Big)^2 + \Big(\frac{d\gamma_2(t)}{dt}\Big)^2} \, dt$$

An analogous relation holds for a continuous, piecewise differentiable curve.

Multiplying the Poisson equation by a sufficiently smooth function v, integrating over Ω and using the first Green formula we find that there holds

$$\int_\Omega f v \, dx = \int_\Omega -\Delta u \, v \, dx = \int_\Omega \nabla u \cdot \nabla v \, dx - \int_\Gamma \partial_n u v \, ds.$$

Now, selecting a space V as

$$V := H_0^1(\Omega) := \{v \in L_2(\Omega);\ \nabla v \in (L_2(\Omega))^2,\ v = 0 \text{ on } \Gamma\},$$

this leads to the formulation

$$u \in V : \quad a(u, v) = (f, v) \quad \forall v \in V \tag{1.3}$$

with

$$a(u, v) := \int_\Omega \nabla u \cdot \nabla v \, dx \quad \text{and} \quad (f, v) := \int_\Omega f v \, dx.$$

Problem (1.3) is called the *variational* or *weak formulation* of (1.1). In this particular case there is an equivalent *minimisation problem*:

$$u \in V : \quad F(u) \le F(v) \quad \forall v \in V \quad \text{where} \quad F(v) := \frac{1}{2} a(v, v) - (f, v). \tag{1.4}$$

Notations and definitions. For the discussion and analysis of (1.3) we need to introduce some definitions and derivatives used for the space $H_0^1(\Omega)$.

Let V be a linear space. $L: V \to \mathbf{R}$ is called a *linear form* if

$$L(\beta v + \theta w) = \beta L(v) + \theta L(w) \quad \forall v, w \in V, \forall \beta, \theta \in \mathbf{R}.$$

$a(\cdot,\cdot)$ is a *bilinear form* on $V \times V$ if $a: V \times V \to \mathbf{R}$ and if it is linear in both arguments:

$$\begin{aligned} a(u, \beta v + \theta w) &= \beta a(u,v) + \theta a(u,w), \\ a(\beta u + \theta v, w) &= \beta a(u,w) + \theta a(v,w) \end{aligned}$$

for all $u, v, w \in V$ and all $\beta, \theta \in \mathbf{R}$. The bilinear form a is called *symmetric* if

$$a(v,w) = a(w,v) \quad \forall v, w \in V.$$

A symmetric bilinear form on $V \times V$ is a *scalar* or *inner product* on V if it is positive definite:

$$a(v,v) > 0 \quad \forall v \in V,\ v \neq 0.$$

Every inner product $\langle \cdot,\cdot \rangle$ on $V \times V$ defines a norm $\|\cdot\|$ on V, and there holds the *Cauchy-Schwarz inequality*

$$|\langle v, w \rangle| \leq \|v\| \, \|w\| \quad \forall v, w \in V. \tag{1.5}$$

Also, remember that a complete normed space with inner product is called a *Hilbert space*.

Now we introduce a weak form of derivatives.

Definition 1.1 *Let $I \subset \mathbf{R}$ be an interval. An element $v \in L_2(I)$ (we call it function) is* weakly differentiable *if there exists $g \in L_2(I)$ such that*

$$\int_I v\phi' \, dx = -\int_I g\phi \, dx \qquad \forall \phi \in C_0^\infty(I).$$

Here, the derivative ϕ' is the classical one. When such a g exists then one defines $v' := g$.

Note that the weak derivative coincides with the classical derivative for a differentiable function. This follows from the integration-by-parts formula. The extension of this definition to higher orders is by induction and to higher dimensions by replacing the above integration-by-parts formula by first Green's formula (cf. Lemma 1.1).

Summary. The boundary value problem (1.1) has the weak formulation (1.3) where $a(\cdot,\cdot)$ is a symmetric bilinear form on $H_0^1(\Omega) \times H_0^1(\Omega)$ (one can prove that it is also positive definite) and where $(f,\cdot)$ is a linear form on $H_0^1(\Omega)$. $L_2(\Omega)$ and

$$H^1(\Omega) := \{v \in L_2(\Omega);\ \nabla v \in (L_2(\Omega))^2\}$$

are Hilbert spaces (derivatives are defined in the weak sense) with inner products and norms

$$\begin{aligned}
(v,w) := (v,w)_{L_2(\Omega)} &:= \int_\Omega vw\,dx, & \|v\|_{L_2(\Omega)} &:= \left(\textstyle\int_\Omega v^2\,dx\right)^{1/2}, \\
(v,w)_{H^1(\Omega)} &:= \textstyle\int_\Omega \left(vw + \nabla v\cdot\nabla w\right)dx, & \|v\|_{H^1(\Omega)} &:= \left(\textstyle\int_\Omega \left(v^2 + |\nabla v|^2\right)dx\right)^{1/2}.
\end{aligned}$$

Moreover, $H_0^1(\Omega)$ provided with the $H^1(\Omega)$-norm is a closed subspace of $H^1(\Omega)$.

The spaces $H^1(\Omega)$ and $H_0^1(\Omega)$ are called *Sobolev spaces.*

Theorem 1.1 *Any solution of* (1.1) *solves* (1.3)*, and the problems* (1.3) *and* (1.4) *are equivalent. Any solution of* (1.3) *which is sufficiently regular solves* (1.1)*.*

Proof. We have already seen that any solution of (1.1) solves (1.3). Now we show that (1.3) and (1.4) are equivalent. Let u solve (1.3) and let $v \in V$. Then set $w = v - u$ so that $v = u + w$ with $w \in V$. We obtain

$$\begin{aligned}
F(v) &= F(u+w) = \frac{1}{2}a(u+w,u+w) - (f,u+w) \\
&= \frac{1}{2}a(u,u) - (f,u) + a(u,w) - (f,w) + \frac{1}{2}a(w,w) \geq F(u)
\end{aligned}$$

since by (1.3), $a(u,w) - (f,w) = 0$ and $a(w,w) \geq 0$. Therefore, u solves (1.4). Now, if u is a solution of (1.4) then for any $v \in V$ and any real number ϵ there holds

$$F(u) \leq F(u + \epsilon v),$$

since $u + \epsilon v \in V$. Therefore, the function g defined by

$$g(\epsilon) := F(u+\epsilon v) = \frac{1}{2}a(u,u) + \epsilon a(u,v) + \frac{\epsilon^2}{2}a(v,v) - (f,u) - \epsilon(f,v)$$

is differentiable, has a minimum at $\epsilon = 0$ and, thus, $g'(0) = 0$. This yields

$$g'(0) = a(u,v) - (f,v) = 0 \qquad \forall v \in V,$$

i.e. u solves (1.3).

Now, to show that a sufficiently smooth solution of (1.3) is also a solution to (1.1) we need that Δu exists and is continuous. Then, considering the property of u that it satisfies

$$\int_\Omega \nabla u \cdot \nabla v\,dx - \int_\Omega f v\,dx = 0 \quad \forall v \in V$$

and integrating by parts, we obtain

$$\int_\Omega (\Delta u + f) v\,dx = 0 \quad \forall v \in V.$$

By the continuity of $\Delta u + f$ this requires that

$$\Delta u + f = 0 \quad \text{pointwise on } \Omega.$$

Since u is continuous, the search for $u \in V = H_0^1(\Omega)$ in particular means that the homogeneous Dirichlet boundary condition is satisfied. This proves that u solves (1.1). $\square$

Exercise 1.1 *Derive the variational formulation and corresponding minimisation problem of the boundary value problem*

$$\begin{aligned} u^{(iv)}(x) &= f(x) \qquad \textit{for } 0 < x < 1, \\ u(0) &= u(1) = u'(0) = u'(1) = 0. \end{aligned}$$

Here, $u^{(iv)}$ denotes the fourth order derivative of u.

1.2 The finite element method for the Poisson equation

The finite element method (FEM) for the solution of (1.1) consists in solving (1.3) or (1.4) within a finite-dimensional subspace V_h of V. This so-called *finite element* or *ansatz space* is constructed by piecewise polynomial functions. The idea is that basis functions have small support. Here we consider the simplest case of continuous piecewise linear functions.

Let us assume, for simplicity, that Ω is a polygonal domain. To define the finite element space we consider a *triangulation* $\mathcal{T}_h = \{K_j : \ j = 1, \ldots, m\}$ of Ω into triangles (or elements) K_j, i.e.

$$\bar{\Omega} = \bigcup_{K \in \mathcal{T}_h} K = K_1 \cup K_2 \cup \ldots \cup K_m.$$

Here we assume that any two triangles are disjoint or intersect at a single vertex or an entire edge. The triangulation $\mathcal{T}_h$ is also called a *mesh* on Ω. With any such mesh we associate the *mesh size* or *mesh width* defined by

$$h = \max_{K \in \mathcal{T}_h} \operatorname{diam}(K) \quad \text{where} \quad \operatorname{diam}(K) := \text{ diameter of } K = \text{ longest side of } K.$$

Our finite element space then is

$$V_h := \{v : \ v \text{ is continuous on } \Omega, \ v|_K \text{ is linear for } K \in \mathcal{T}_h, \ v = 0 \text{ on } \Gamma\}.$$

The *finite element method* for (1.1) reads:

$$\text{find } u_h \in V_h \text{ such that} \quad F(u_h) \leq F(v) \qquad \forall v \in V_h \tag{1.6}$$

in the form of a minimisation problem, or

$$\text{find } u_h \in V_h \text{ such that} \quad a(u_h, v) = (f, v) \qquad \forall v \in V_h \tag{1.7}$$

in discrete variational form. Of course, as in the proof of Theorem 1.1 one sees that (1.6) and (1.7) are equivalent. Historically, (1.6) is called the *Ritz method* and (1.7) the *Galerkin method*.

To calculate u_h (theoretically or on a computer) one transforms the discrete problem (i.e. (1.6) or (1.7)) into a system of linear equations.

One can identify any element of V_h by its values at the *nodes* N_j $(j = 1, \ldots, M)$ of the mesh (the set of vertices of the triangles). In particular, the dimension of V_h is the number M of interior nodes of the mesh $\mathcal{T}_h$ (the values on boundary nodes, the ones on Γ, are fixed by definition of V_h). It is immediate that the functions $\varphi_j \in V_h$ defined by

$$\varphi_j(N_i) = \delta_{ij} \equiv \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \qquad i, j = 1, \ldots, M$$

form a basis of V_h (see Figure 1.1), they are called *basis functions*.

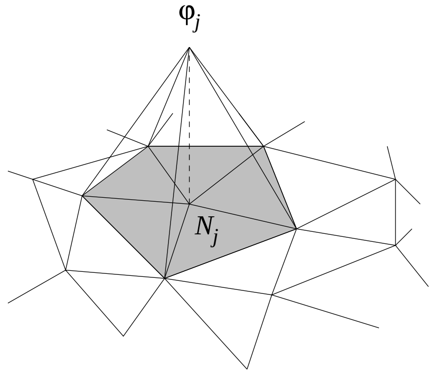

Figure 1.1: Piecewise linear basis function φ_j.

The support of φ_j consists of all elements that have N_j as a vertex. Note that this number of elements depends on the mesh construction and can be different for different nodes. One can represent any $v \in V_h$ as a linear combination of the basis functions,

$$v = \sum_{j=1}^{M} \eta_j \varphi_j \quad \text{where} \quad \eta_j = v(N_j).$$

In particular, the finite element approximation u_h has the unique representation

$$u_h(x) = \sum_{i=1}^{M} \xi_i \varphi_i(x), \qquad \xi_i = u_h(x_i) \tag{1.8}$$

and it is enough to determine $\xi = (\xi_1, \ldots, \xi_M) \in \mathbf{R}^M$.

The following lemma is immediate.

Lemma 1.2 *The solution u_h of* (1.7) *is given by* (1.8) *where ξ is the solution of the linear system*

$$A\xi = b \tag{1.9}$$

where $A = (a_{ij})$ is the $M \times M$ stiffness matrix *with elements*

$$a_{ij} = a(\varphi_i, \varphi_j) = \int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \, dx, \quad i, j = 1, \ldots, M,$$

and $b = (b_i) \in \mathbf{R}^M$ is the load vector *with*

$$b_i = (f, \varphi_i) = \int_\Omega f\varphi_i \, dx, \quad i = 1, \ldots, M.$$

1.2.1 Properties and assembly of the stiffness matrix

The stiffness matrix A of (1.9) is symmetric and positive definite:

$$\eta \cdot A\eta > 0 \qquad \forall \eta \in R^M, \eta \neq 0.$$

This follows from the symmetry and positive definiteness of the bilinear form $a(\cdot, \cdot)$.

The symmetry and positive definiteness of A are important properties when solving the linear system (1.9). For moderate dimensions M it can be solved by the *Cholesky method*, and large systems can be solved iteratively by the *conjugate gradient method* (CG-method). Both methods are the most efficient ones in their class (of direct and iterative methods, respectively) and require symmetric, positive definite matrices.

Another property of A is that it has only a few non-zero elements, it is a *sparse* matrix. Indeed, whenever two basis functions φ_i, φ_j are associated with nodes of different triangles then the measure of the intersection of the supports of φ_i and φ_j is zero so that $a_{ij} = a(\varphi_i, \varphi_j) = 0$. For large numbers of unknowns M the number of non-zero elements of A grows only linearly in M (whereas there are M^2 entries of A in total). This fact, and the special structure of A, can be used to solve the linear system efficiently by only storing $O(M)$ numbers. (Here, $O(M)$ denotes a number that grows at most linearly in M when $M \to \infty$.)

To *assemble* the stiffness matrix one uses an element-oriented strategy. Using the decomposition $\bar{\Omega} = \cup_{K \in \mathcal{T}_h} K$ we find for any $i, j \in \{1, \ldots, M\}$ that

$$a(\varphi_i, \varphi_j) = \int_\Omega \nabla\varphi_i \cdot \nabla\varphi_j \, dx = \sum_{K \in \mathcal{T}_h} \int_K \nabla\varphi_i \cdot \nabla\varphi_j \, dx =: \sum_{K \in \mathcal{T}_h} a_K(\varphi_i, \varphi_j). \tag{1.10}$$

There holds $a_K(\varphi_i, \varphi_j) = 0$ unless both nodes N_i and N_j are vertices of the triangle K. Therefore, to calculate $a_K(\varphi_i, \varphi_j)$, one only needs to consider the numbers $i, j \in \{1, \ldots, M\}$ which correspond to nodes N_i, N_j of K. For arbitrary (but fixed) $K \in \mathcal{T}_h$ let N_i, N_j, N_k denote its three vertices. We then call the 3×3-matrix

$$A_K := \begin{pmatrix} a_K(\varphi_i, \varphi_i) & a_K(\varphi_i, \varphi_j) & a_K(\varphi_i, \varphi_k) \\ & a_K(\varphi_j, \varphi_j) & a_K(\varphi_j, \varphi_k) \\ \text{sym} & & a_K(\varphi_k, \varphi_k) \end{pmatrix} \tag{1.11}$$

the *element* or *local stiffness matrix* for K. In order to calculate the stiffness matrix A one calculates all the element stiffness matrices A_K and then forms A by using (1.10). This process is called the *assembly* of A. A is sometimes called *global stiffness matrix* to distinguish it from the local matrices. An analogous procedure is used to construct the load vector b.

To calculate A_K one obviously needs only the restrictions of the basis functions φ_i, φ_j, φ_k onto K. Let us denote these restrictions by

$$\psi_i := \varphi_i|_K, \quad \psi_j := \varphi_j|_K, \quad \psi_k := \varphi_k|_K.$$

Each of these three functions is linear (on K) and has the value 1 at exactly one vertex and vanishes at the other two vertices. Any linear function w on K can be represented by

$$w = w(N_i)\psi_i + w(N_j)\psi_j + w(N_k)\psi_k.$$

The functions ψ_i, ψ_j, ψ_k are called *local* or *element basis functions* on K.

Exercise 1.2 *Consider the triangle $\tilde{K}$ with vertices $\tilde{N}_1 = (0,0)$, $\tilde{N}_2 = (h,0)$ and $\tilde{N}_3 = (0,h)$. Define the local (linear) basis functions associated with the vertices and show that the local stiffness matrix for $\tilde{K}$ is given by*

$$\tilde{A} = (\tilde{a}_{ij})_{i,j=1}^3 = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}.$$

Also, convince yourself that a translation or rotation of $\tilde{K}$ does not alter this matrix.

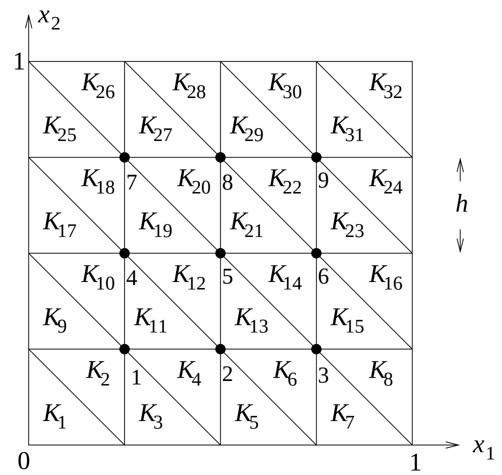

Figure 1.2: Uniform triangulation with $h = 1/4$ for Example 1.1.

Example 1.1 *Let us consider a square Ω with side length 1 and let $\mathcal{T}_h$ be a uniform triangulation of Ω with $h = 1/4$, see Figure 1.2. (Here, for simplicity, h denotes the smallest side length of the triangles which is proportional to their diameter since they are shape regular.) The nodes N_i appear as numbers $i = 1, \ldots, 9$ and the elements are K_i, $i = 1, \ldots, 32$. We use the local stiffness matrix $\tilde{A} = (\tilde{a}_{ij})$ from Exercise 1.2 and formula (1.10) to assemble the global stiffness matrix. For instance, noting that the supports of φ_4, $\varphi_1\varphi_4$ and $\varphi_2\varphi_4$ are $\cup_{i\in\{10,11,12,19,18,17\}} K_i$, $K_{10} \cup K_{11}$ and $K_{11} \cup K_{12}$, respectively, we obtain*

$$\begin{aligned} a_{4,4} = \sum_{i\in\{10,11,12,19,18,17\}} a_{K_i}(\varphi_4, \varphi_4) &= \tilde{a}_{1,1} + \tilde{a}_{3,3} + \tilde{a}_{2,2} + \tilde{a}_{1,1} + \tilde{a}_{3,3} + \tilde{a}_{2,2} \\ &= 1 + 1/2 + 1/2 + 1 + 1/2 + 1/2 = 4, \end{aligned}$$

$$\begin{aligned} a_{1,4} &= \sum_{i\in\{10,11\}} a_{K_i}(\varphi_1, \varphi_4) = \tilde{a}_{3,1} + \tilde{a}_{1,3} = -1/2 - 1/2 = -1, \\ a_{2,4} &= \sum_{i\in\{11,12\}} a_{K_i}(\varphi_1, \varphi_4) = \tilde{a}_{2,3} + \tilde{a}_{3,2} = 0 + 0 = 0. \end{aligned}$$

Proceeding in this way we obtain the global stiffness matrix

$$A = \begin{pmatrix} 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 4 & -1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & -1 & 4 & 0 & 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 4 & -1 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & -1 & 4 & -1 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & -1 & 4 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 & 0 & 4 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & -1 & 4 \end{pmatrix}.$$

Exercise 1.3 *Consider the situation described in Example 1.1 but with $h = 1/3$ instead of $h = 1/4$. For right-hand side function $f(x) = 1$ ($x \in \Omega$) assemble the linear system (1.9), determine the solution u_h of (1.7) and calculate $u_h(1/2, 1/2)$.*

1.3 Galerkin orthogonality

Any variational formulation

$$u \in V : \quad a(u, v) = L(v) \quad \forall v \in V \tag{1.12}$$

with corresponding finite element scheme

$$u_h \in V_h \subset V : \quad a(u_h, v) = L(v) \quad \forall v \in V_h$$

translates into the *Galerkin orthogonality*

$$a(u - u_h, v) = 0 \quad \forall v \in V_h. \tag{1.13}$$

Let us consider the following variant of the homogeneous Dirichlet problem:

$$\begin{aligned} -\Delta u + u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \Gamma. \end{aligned} \tag{1.14}$$

The corresponding variational formulation is

$$u \in H_0^1(\Omega) : \quad (\nabla u, \nabla v) + (u, v) = (f, v) \quad \forall v \in H_0^1(\Omega), \tag{1.15}$$

and can also be written in the general form (1.12) with $V = H_0^1(\Omega)$, $a(u, v) = (\nabla u, \nabla v) + (u, v)$ and $L(v) = (f, v)$. We note that in fact $a(u, v) = \langle u, v\rangle$ is the standard inner product in $H_0^1(\Omega)$ such that the variational formulation renders like

$$u \in H_0^1(\Omega) : \quad \langle u, v\rangle = L(v) \quad \forall v \in H_0^1(\Omega).$$

Introducing a finite element space $V_h \subset H_0^1(\Omega)$ one has a corresponding finite element scheme and the Galerkin orthogonality (1.13) becomes

$$\langle u - u_h, v\rangle = 0 \quad \forall v \in V_h. \tag{1.16}$$

The relation (1.16) means that the finite element error $u - u_h$ is orthogonal to the finite element space V_h. In particular, u_h is the *projection* with respect to the inner product $\langle \cdot, \cdot\rangle$ of u onto V_h. Figure 1.3 gives a geometric description of this fact for the case $V = \mathbf{R}^2$ with Euclidean inner product and a one-dimensional subspace $V_h \subset V$. This property proves the following best approximation property:

$$\|u - u_h\|_{H^1(\Omega)} \leq \|u - v\|_{H^1(\Omega)} \quad \forall v \in V_h. \tag{1.17}$$

Note that with the finite element method for (1.14) we calculate the projection of the exact solution u onto V_h without actually knowing it. It only requires the solution of a sparse linear system $A\xi = b$ with symmetric, positive definite matrix A.

Exercise 1.4 *Show that* (1.16) *and* (1.17) *are equivalent.*

1.4 Natural and essential boundary conditions

So far we have considered only Dirichlet boundary conditions where the sought solution is prescribed on the boundary of the domain. There is another important type of boundary conditions where the normal derivative is prescribed. Such a problem is called *Neumann problem*:

$$\begin{aligned} -\Delta u + u &= f \qquad \text{in } \Omega, \\ \partial_n u &= g \qquad \text{on } \Gamma. \end{aligned} \tag{1.18}$$

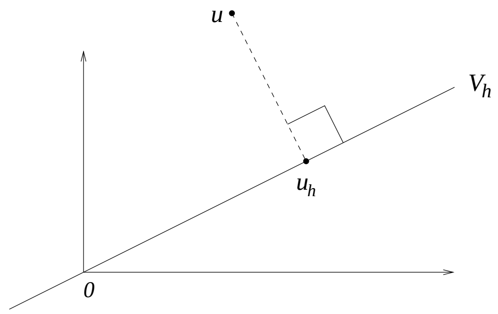

Figure 1.3: Projection of u onto V_h.

Here, f and g are given functions and $\partial_n u$ denotes, as introduced before, the outward normal derivative of u on Γ. The boundary condition is called *Neumann boundary condition.*

The variational formulation of (1.18) is

$$u \in H^1(\Omega): \quad a(u,v) = (f,v) + (g,v)_\Gamma \quad \forall v \in H^1(\Omega), \tag{1.19}$$

where

$$a(u,v) := (\nabla u, \nabla v) + (u,v) \quad \text{and} \quad (g,v)_\Gamma := \int_\Gamma gv\, ds.$$

Correspondingly, the minimisation problem is

$$u \in H^1(\Omega): \quad F(u) \leq F(v) \quad \forall v \in H^1(\Omega) \tag{1.20}$$

where

$$F(v) := \frac{1}{2} a(v,v) - (f,v) - (g,v)_\Gamma.$$

Theorem 1.2 *Any solution u of* (1.18) *solves* (1.19)*. If u is a sufficiently regular solution of* (1.19) *then it solves* (1.18)*. Moreover, problems* (1.19) *and* (1.20) *are equivalent.*

Proof. The equivalence of (1.19) and (1.20) is analogous to the situation in Theorem 1.1. Now assume that u solves (1.18). We multiply the differential equation in (1.18) by a test function $v \in H^1(\Omega)$ and integrate over Ω. Using that $\partial_n u = g$ on Γ, the first Green formula (Lemma 1.1) gives

$$\begin{aligned} (f,v) &= \int_\Omega (-\Delta u + u)\, v\, dx = -\int_\Gamma \partial_n u v\, ds + \int_\Omega \nabla u \cdot \nabla v\, dx + \int_\Omega uv\, dx \\ &= -(g,v)_\Gamma + (\nabla u, \nabla v) + (u,v) = a(u,v) - (g,v)_\Gamma. \end{aligned}$$

This is (1.19). Now let u be a sufficiently smooth function that solves (1.19). Using again Green's first formula we obtain

$$(f,v) + (g,v)_\Gamma = a(u,v) = \int_\Gamma \partial_n u v\, ds + \int_\Omega (-\Delta u + u)\, v\, dx,$$

that is

$$\int_\Omega (-\Delta u + u - f)\, v\, dx + \int_\Gamma (\partial_n u - g)\, v\, ds = 0 \quad \forall v \in H^1(\Omega). \tag{1.21}$$

In particular, (1.21) holds for any $v \in H^1(\Omega)$ with $v = 0$ on Γ, that is

$$\int_\Omega (-\Delta u + u - f)\, v\, dx = 0 \quad \forall v \in H_0^1(\Omega).$$

For sufficiently smooth u this is only possible if

$$-\Delta u + u - f = 0 \quad \text{in } \Omega.$$

Taking this relation (it is the wanted differential equation) into account (1.21) becomes

$$\int_\Gamma (\partial_n u - g)\, v\, ds = 0 \quad \forall v \in H^1(\Omega).$$

By varying the test function $v \in H^1(\Omega)$ appropriately it can be seen that this requires

$$\partial_n u - g = 0 \quad \text{on } \Gamma.$$

This finishes the proof of the theorem. □

Remark 1.2 *Note that the Neumann boundary condition appears in the variational formulation (via the linear form on the right-hand side) and is not incorporated in the space $V = H^1(\Omega)$. It is therefore called* natural boundary condition. *In contrast, a Dirichlet boundary condition of the type $u = 0$ on Γ enters the variational formulation by choosing V appropriately to reflect this condition, $V = H_0^1(\Omega) \subset H^1(\Omega)$ in this case. Therefore, Dirichlet boundary conditions are also called* essential boundary conditions. *This difference in incorporating boundary conditions is inherited by the finite element schemes.*

To define a finite element scheme for the approximate solution of (1.19) we choose a finite-dimensional subspace $V_h \subset H^1(\Omega)$. To this end we consider as before a mesh $\mathcal{T}_h$ consisting of triangles K. The simplest choice of V_h is

$$V_h := \{v :\ v \text{ is continuous on } \Omega,\ v|_K \text{ is linear } \forall K \in \mathcal{T}_h\}.$$

Note that we do not ask $v \in V_h$ to vanish on Γ. All the nodes of $\mathcal{T}_h$ including the ones on Γ are now being taken into account. The finite element method then is:

$$\text{find } u_h \in V_h : \quad a(u_h, v) = (f, v) + (g, v)_\Gamma \quad \forall v \in V_h. \tag{1.22}$$

Note that u_h in general does not satisfy the Neumann boundary condition. One can rather show that $\partial_n u_h \to g$ $(h \to 0)$ in an appropriate norm.

Exercise 1.5 *Let* $\Omega \subset \mathbf{R}^2$ *be a bounded domain with Lipschitz continuous boundary* Γ *and let* Γ *be decomposed into two non-empty curves* Γ_1 *and* Γ_2*:* $\Gamma = \bar{\Gamma}_1 \cup \bar{\Gamma}_2$ *and* $\Gamma_1 \cap \Gamma_2 = \emptyset$*. For sufficiently smooth functions* f *and* g *give a variational formulation of the* mixed boundary value problem

$$\begin{aligned} -\Delta u + u &= f \qquad \textit{in } \Omega, \\ u &= 0 \qquad \textit{on } \Gamma_1, \\ \partial_n u &= g \qquad \textit{on } \Gamma_2. \end{aligned}$$