Computational PDEs I+II

(Theorie und Numerik partieller Differentialgleichungen I+II)

Lectures at the Universität Jena

Last update: 25th April 2025

D. Gallistl

Contents

CHAPTER 1

Standard Finite Element Methods

1.1. Tools

1.1.1. Elementary Hilbert space theory. Hilbert spaces are taught in detail in any class on linear functional analysis. Here we only focus on some very basic properties.
Let X be a (real) linear space. Given a symmetric and positive definite bilinear form $(\cdot,\cdot)_X$, we define $\|x\|_X = \sqrt{(x,x)_X}$ for any $x \in X$. It is elementary to establish the Cauchy–Schwarz inequality

$$(x,y)_X \leq \|x\|_X \, \|y\|_X \quad \text{for any } x, y \in X.$$

It can be shown that $\|\cdot\|_X$ defines a norm on X (thereby justifying the notation). The proofs are the same as in the case of Euclidean vector spaces and left as an exercise. The difference to Euclidean spaces is that X can be infinite-dimensional and therefore need not be complete. If it is complete, we call X a Hilbert space.

Definition 1.1 (Hilbert space). A linear space X (over $\mathbb{R}$) equipped with a symmertic and positive definite bilinear form $(\cdot,\cdot)_X$ is called *Hilbert space* if it is complete with respect to the norm $\|\cdot\|_X := \sqrt{(\cdot,\cdot)_X}$. ♦

Basically, Hilbert spaces are Banach spaces with an Euclidean structure.

Lemma 1.2 (parallelogram law). *In a Hilbert space X, every $(a,b) \in X^2$ satisfies*

$$\|a-b\|_X^2 + \|a+b\|_X^2 = 2(\|a\|_X^2 + \|b\|_X^2).$$

Proof. If we expand both terms on the left-hand side with the binomial identity, we see that the mixed terms cancel. What remains are the terms on the right-hand side. □

Theorem 1.3 (projection on complete subspaces). *Let X be a Hilbert space with inner product $(\cdot,\cdot)_X$ and let $Y \subseteq X$ be a complete linear subspace. Given $x \in X$, there exists a unique element $Px \in Y$ with the property*

$$\|x - Px\|_X = \inf_{y \in Y} \|x - y\|_X.$$

The element Px is unique and characterized by the property

$$(x - Px, y)_X = 0 \quad \textit{for all } y \in Y.$$

Proof. We abbreviate $\delta := \inf_{y\in Y} \|x-y\|_X$ Let $(y_k)_k$ be a sequence in Y with $\|x - y_k\|_X \to \delta$ as $k \to \infty$. To prove that the sequence is Cauchy, we let $m, n \geq 0$ and choose $a = x - y_m$, $b = x - y_n$ in the parallelogram law, which results in

$$\|y_m - y_n\|_X^2 + 4\|x - \frac{1}{2}(y_m + y_n)\|_X^2 = 2(\|x - y_m\|_X^2 + \|x - y_n\|_X^2).$$

Since y_m, y_n are from the linear space Y, their average lies in Y, and the second term on the left-hand side is bounded from below by 4δ. Since the right-hand side converges to the same value, we deduce $\|y_m - y_n\|_X \to 0$ as $m, n \to \infty$ so that $(y_k)_k$ is a Cauchy sequence. Since Y is complete, the sequence has a limit, denoted by y, which lies in Y and satisfies $\|x - y\| = \delta$. For proving uniqueness, we

assume that there are $y, y' \in Y$ realizing the infimum. The above argument with the parallelogram law applied to y, y' instead of y_m, y_n shows that $y = y'$, which proves uniqueness. We thus denote $Px := y$.
For an arbitrary $z \in Y$ and $\varepsilon \in [0,1]$, the convex combination $(1-\varepsilon)Px+\varepsilon z$ belongs to Y, so that we infer with elementary manipluations

$$\|x - Px\|_X^2 = \delta^2 \leq \|x - (1-\varepsilon)Px - \varepsilon z\|_X^2 = \|(x - Px) - \varepsilon(z - Px)\|_X^2.$$

Expanding the right-hand side results in the estimate

$$\|x - Px\|_X^2 \leq \|x - Px\|_X^2 + \varepsilon^2\|z - Px\|_X^2 + 2\varepsilon(x - Px, z - Px)_X.$$

Simplifying, dividing by ε, and letting $\varepsilon \to 0$, we see with the substitution $y := z - Px$ that $0 \leq (x - Px, y)_X$ for all $y \in Y$. Since this must be true for $\pm y$, the bilinearity proves the asserted variational identity. □

Definition 1.4. The map $P : X \to Y$ from Theorem 1.3 is called *orthogonal projection* to Y. ♦

It is easy to see that the orthogonal projection P to a subspace Y is linear and nonexpansive, that is $\|P\|_{L(X,Y)} \leq 1$, see the exercises.
We recall the dual space $X^* := L(X, \mathbb{R})$, which is the space of continuous linear functionals over X. The Riesz representation theorem states that there exists an isometric isomorphism between X and X^*. The proof is taught in every course on linear functional analysis and we will briefly discuss the proof in what follows.

Theorem 1.5 (Riesz representation theorem). *Let X be a Hilbert space with inner product $(\cdot,\cdot)_X$ and let $F \in X^*$ be a continuous linear functional. Then there exists a unique element $x \in X$ with the property*

$$(y, x)_X = F(y) \quad \text{for all } y \in X.$$

The dependence of x on F is linear and the element x satisfies $\|x\|_X = \|F\|_{X^*}$.

Proof. We consider the map $J : X \to X^*$ defined by

$$x \mapsto J(x) = [y \mapsto (y, x)_X]$$

or $J(x) = (\cdot, x)_X$ for short. It is direct to check that J is linear and satisfies $\|x\|_X \leq \|J(x)\|_{X^*} \leq \|x\|_X$ so that it is injective and an isometry. We are left with showing that J is surjective. Given some nonzero $F \in X^*$, we denote by P the orthogonal projection to the kernel $\ker(F)$ (which is closed and thus complete, see the exercises). We choose $b \in X$ with $F(b) = 1$ and set $y := b - Pb$. By scaling the element y, we can now decompose any $z \in X$ in a part in the kernel of F and a multiple of y, namely

$$z = (z - F(z)y) + F(z)y.$$

By construction, y is orthogonal to any element of $\ker(F)$, so that we compute

$$(z, y)_X = (z - F(z)y, y)_X + (F(z)y, y)_X = (F(z)y, y)_X = F(z)\|y\|_X^2.$$

Rearranging this formula reveals

$$F(z) = \|y\|_X^{-2}(z, y)_X = J(\|y\|_X^{-2}y)(z) \quad \text{for all } z \in X$$

(note that y is nonzero because $F(y) = 1$). We thus have shown $F = J(x)$ for $x = \|y\|_X^{-2}y$, whence J is surjective.

□

1.2. Linear elliptic problems

1.2.1. Dirichlet problem and Dirichlet principle. In this lecture we study partial differential equations (PDEs) and their numerical approximation. We confine ourselves to linear equations of second order. Let us first define what we mean by this.

DEFINITION 1.6. Let $n \in \mathbb{N}$ and $\Omega \subseteq \mathbb{R}^n$ be an open subset. Let furthermore a map $F : \mathbb{R}^{n\times n} \times \mathbb{R}^n \times \mathbb{R} \times \Omega \to \mathbb{R}$ be given. We call the equation

$$F(D^2u(x), \nabla u(x), u(x), x) = 0 \quad \text{for all } x \in \Omega$$

a *partial differential equation of 2nd order*. Any function $u : \Omega \to \mathbb{R}$ satisfying the above relation is called a *solution*. ♦

The foregoing definition is rather abstract. At the same time, it implicitly requires further properties (differentiability) of the solution, which are not stated explicitly. We will work with this basic definition and will proceed with examples. The equation is called *partial* differential equation because it involves partial derivatives of the solution (in contrast to *ordinary differential equations (ODEs)*, which only depend on one scalar variable. The notion of *2nd order* describes that the highest involved derivative of u has order 2. At this point, the function F can be arbitrarily nonlinear.

EXAMPLE 1.7. For a given function $f \in C(\Omega)$ (usually referred to as *right-hand side*) and F given by $F(A, b, c, x) = \det A - f(x)$, we obtain the equation

$$\det D^2u(x) = f(x).$$

It is called *Monge–Ampère equation*.

Recall the Laplacian $\Delta u(x) = \operatorname{div} \nabla u(x) = \sum_{j=1}^n \partial_{jj} u(x) = \operatorname{tr} D^2u(x)$, where $\operatorname{tr} A$ denotes the trace of a matrix A.

EXAMPLE 1.8. For $f \in C(\Omega)$ and $F(A, b, c, x) = \operatorname{tr} A + f(x)$ we obtain *Poisson's equation*

$$-\Delta u(x) = f(x).$$

What is the difference between these two examples? Poisson's equation is *linear*. This means that, given solutions u to the right-hand side f and v to the right-hand side g, the equation

$$-\Delta w(x) = \alpha f(x) + \beta g(x)$$

will be satisfied by the linear combination $w := \alpha u + \beta v$, $(\alpha, \beta \in \mathbb{R})$. This is easy to verify. It is also elementary to verify that the Monge–Ampère equation does not have this property. We expect in general that

$$\det D^2(u(x) + v(x)) \neq f(x) + g(x),$$

for solutions u and v to right-hand sides f and g, respectively. Convince yourself of this fact by setting up suitable examples.

DEFINITION 1.9. A 2nd order PDE is called *linear*, if it is of the form

$$\sum_{|\alpha|\leq 2} a_\alpha(x)\partial^\alpha u(x) = f(x).$$

Here, a_α and f are given functions over Ω. The above sum runs over all multi-indices α of length ≤ 2, and ∂^α is the partial derivative with respect to α. ♦

We will start this lecture by considering Poisson's equation, the most basic instance of a PDE that is rich enough to highlight all relevant concepts. Generally, we pose the questions of *existence* of a solution to a PDE and its *uniqueness*. Even for Poisson's equation we will quickly reach certain limits that we will later overcome with tools of linear functional analysis.
Clearly, solutions to Poisson's equation are not unique without any further constraints being imposed. For instance, any solution can be shifted by an arbitrary affine function and will still remain a solution. We will thus consider the *Dirichlet problem*, which imposes a zero boundary condition on the solution. This PDE is posed on a domain $\Omega \subseteq \mathbb{R}^n$ which is open, bounded, and connected.

DEFINITION 1.10. Let $\Omega \subseteq \mathbb{R}^n$ be open, bounded, and connected. A function $u \in C^2(\Omega) \cap C(\bar{\Omega})$ is said to solve the Dirichlet problem with right-hand side $f \in C(\Omega)$ if it satisfies

$$-\Delta u = f \text{ in } \Omega \quad \text{und} \quad u = 0 \text{ on } \partial\Omega.$$

♦

It will often be important to impose more structure on the boundary $\partial\Omega$.

DEFINITION 1.11. A domain Ω has a *Lipschitz boundary*, if there are finitely many open sets $U^1, \dots, U^N \subseteq \mathbb{R}^n$ that cover a neighbourhood of the boundary $\partial\Omega$ and have the property that, for any $j \in \{1, \dots, N\}$, the set $\partial\Omega \cup U^j$ can be represented as the graph of a Lipschitz function so that γ_j such that $\Omega \cap U^j$ lies on exactly one side of the graph.
Let us give a formal definition in two dimensions. There exists some $N \in \mathbb{N}$, finitely many open sets $U^1, \dots, U^N$, open intervals $I_j \subset \mathbb{R}$ and Lipschitz continuous functions $\gamma_j : \bar{I}_j \to \mathbb{R}^2$ on $\bar{I}_j$ with the following property: The U^j cover a neighbourhood U of $\partial\Omega$, i.e. $U \subseteq \cup_{j=1}^N U^j$, and any U^j satisfies (after some shift and rotation of the coordinate system)

- $U^j \cap \partial\Omega = \{(y, \gamma(y)) : y \in I\}$
- $U^j \cap \Omega \subseteq \{(y, z) : y \in I, z > \gamma(y)\}$

♦

In our examples we will mostly deal with simple Lipschitz domains with boundaries consisting of piecewise smooth curve segments. It is known that Lipschitz domains posses, almost everywhere on the boundary, a well-defined outer unit normal vector ν. The divergence theorem teaches us the following: For a bounded Lipschitz domain Ω and a vector field $v \in C^1(\Omega; \mathbb{R}^n)$ we have

$$\int_{\partial\Omega} v \cdot \nu \, ds = \int_\Omega \operatorname{div} v \, dx.$$

Here (and throughout this text) we denote integration with respect to the n-dimensional Lebesgue measure by the symbol "dx" while integration with respect to the $(n-1)$-dimensional surface measure is indicated by "ds". The divergence theorem implies the formula of integration by parts: For two differentiable functions u and v we have

$$\int_\Omega (u \, \partial_j v + v \, \partial_j u) dx = \int_{\partial\Omega} uv \, \nu_j ds$$

for any $j \in \{1, \dots, n\}$, where ν_j is the jth component of the outer unit normal. A variant thereof is called *Green's formula*

$$\int_\Omega (u\Delta v + \nabla u \cdot \nabla v) dx = \int_{\partial\Omega} u \frac{\partial v}{\partial \nu} ds,$$

where $v \in C^2(\Omega) \cap C^1(\bar{\Omega})$ is assumed.

THEOREM 1.12 (uniqueness). *The solution to the Dirichlet problem is unique.*

PROOF. Let u and v be two solutions to the Dirichlet problem with right-hand side f. The linearity then implies that the difference $w := u - v$ satisfies $-\Delta w = 0$. Integration by parts results in

$$0 = -\int_\Omega w \Delta w \, dx = \int_\Omega |\nabla w|^2 \, dx.$$

Here, no boundary term occurs because w vanishes on the boundary. Therefore, ∇w equals zero almost everywhere in Ω. Hence, w is constant; and since $w = 0$ on $\partial\Omega$, we have that w is the constant zero function whence $u = v$ in Ω. □

At this point we are not yet in the position to formulate a satisfactory existence theory. With the spaces of differentiable functions used above we may lose control over derivatives, which makes it often difficult to justify numerical methods. Let us discuss the following illustrative example.

EXAMPLE 1.13. Let $\Omega = (-1,1)^2 \setminus ([0,1] \times [-1,0])$ be the Γ-shaped (or L-shaped) domain. Let u be given by

$$u(x,y) = (1-x^2)(1-y^2) r^{2/3} \sin\left(\frac{2\varphi}{3}\right).$$

Here, we use polar coordinates $0 < r < 1$ and $0 < \varphi < 3\pi/2$; note that $x = r\cos\varphi$ and $y = r\sin\varphi$. One can verify that u satisfies $-\Delta u = f$ for some $f \in C^0(\bar\Omega)$ and $u|_{\partial\Omega} = 0$. But u does not possess bounded derivatives and, thus, does not belong to $C^1(\bar\Omega)$.

We will now characterize the Dirichlet problem as an optimization problem. To this end, we will employ basic methods from the calculus of variations. The most important tool is the following.

LEMMA 1.14 (fundamental lemma of calculus of variations).

(a) *Let the function $g \in C^0(\Omega)$ satisfy*

$$\int_\Omega g\psi \, dx = 0$$

for all $\psi \in C_c^\infty(\Omega)$ (smooth functions with compact support). Then $g = 0$ holds in the whole domain Ω.

(b) *The assertion of (a) remains valid if $g \in L^1_{\mathrm{loc}}(\Omega)$ (with the same conclusion almost everywhere in Ω).*

PROOF. Exercise. □

THEOREM 1.15 (Dirichlet principle). *Let $\Omega \subseteq \mathbb{R}^n$ be a bounded Lipschitz domain. Let $V = \{v \in C^1(\bar\Omega) : v = 0 \text{ on } \partial\Omega\}$ and $f \in C^0(\Omega)$. A function $u \in V$ satisfies $-\Delta u = f$ in Ω if and only if it minimizes the functional $J : V \to \mathbb{R}$ given by*

$$J(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 dx - \int_\Omega f v \, dx \qquad (v \in V)$$

over V. Here, $|\cdot|$ denotes the Euclidean norm. Any solution to the Dirichlet problem in particular satisfies the necessary condition

$$\int_\Omega \nabla u \cdot \nabla v dx = \int_\Omega f v dx \quad \text{for all } v \in V.$$

Note that above we chose $V \subseteq C^1(\bar\Omega)$ to ensure that the integral involving the gradient is finite. One can weaken this requirement.

PROOF. If $u \in V$ minimizes the functional J, then a necessary criterion is that $J \leq J(u+tv)$ for small perturbations $t > 0$, $v \in V$. This means that $J(u+tv)$ has a minimum at $t = 0$, which implies for the directional derivative that

$$0 = \frac{d}{dt} J(u+tv).$$

The chain rule implies

$$\left(\frac{d}{dt} \frac{1}{2} \int_\Omega |\nabla(u+tv)|^2 dx \right) \Big|_{t=0} = \int_\Omega \nabla u \cdot \nabla v dx.$$

It is furthermore elementary to verify

$$\left(\frac{d}{dt} \int_\Omega f\,(u+tv)\,dx \right) \Big|_{t=0} = \int_\Omega f v dx.$$

This shows that necessary condition

$$\int_\Omega \nabla u \cdot \nabla v dx = \int_\Omega f v dx \quad \text{for all } v \in V. \tag{1}$$

Furthermore, integration by parts (Green's formula) implies

$$\int_\Omega \nabla u \cdot \nabla v dx = \int_\Omega (-\Delta u) v\, dx.$$

Due to the fact that $v \in V$, no boundary term occurs. Altogether, we have shown that

$$\int_\Omega (-\Delta u - f) v\, dx = 0 \quad \text{for all } v \in V.$$

We thus conclude with the fundamental lemma of calculus of variations that $-\Delta u = f$ holds pointwise in Ω.

Let us now assume that $u \in V$ solves Poisson's equation. Green's formula then implies (1) for every $v \in V$. A direct computation, with arbitrary $v \in V$, results in

$$\begin{aligned} J(u+v) - J(u) &= \frac{1}{2} \int_\Omega \left(|\nabla(u+v)|^2 - |\nabla u|^2 \right) dx - \int_\Omega f v\, dx \\ &= \int_\Omega \nabla u \cdot \nabla v\, dx - \int_\Omega f v\, dx + \frac{1}{2} \int_\Omega |\nabla v|^2\, dx \\ &= \frac{1}{2} \int_\Omega |\nabla v|^2\, dx \geq 0. \end{aligned}$$

For the first identity we have used the elementary formula $|b|^2 - |a|^2 = 2a \cdot (b-a) + |a-b|^2$; for the second identity we have used relation (1). Thus, J is minimal at u. □

Dirichlet's principle shows that solving Poisson's equation with boundary conditions is equivalent to solving the corresponding minimization problem. In terms of calculus of variations we say that Poisson's equation is the *Euler–Lagrange equation* corresponding to the minimization problem. Thus, we ask the question under which conditions and in which spaces minimizers of the functional J exist. This will lead us (later in this course) to the concept of *Sobolev spaces*. For the moment we just remark that the formulation as a minimization problem requires weaker conditions on u than the original Dirichlet problem: We only need *first derivatives* to exist. The Laplacian does not explicitly show up in the functional J. It is contained implicitly or *weakly* in that formulation. We will formalize this via the concept of *weak derivatives*.

1.2.2. Weak derivatives and discrete functions. The Dirichlet principle showed us that we can understand derivatives in some weaker sense.

DEFINITION 1.16 (weak derivative). Let $\Omega \subseteq \mathbb{R}^n$ be open. Let $v \in L^1_{\text{loc}}(\Omega)$ and $j \in \{1, \dots, n\}$. If there exists a function $g \in L^1_{\text{loc}}(\Omega)$ with the property

$$\int_\Omega v \partial_j \psi \, dx = -\int_\Omega g\psi \, dx \quad \text{for all } \psi \in C_c^\infty(\Omega),$$

then this function g is called the *weak partial derivative* of v with respect to the direction j, and it is denoted by $\partial_j v$. The vector of all partial derivatives is denoted (provided it exists) by ∇v. ♦

REMARK 1.17. The weak derivative is unique (see problems). ♦

The idea behind this definition is to extend the common notion of differentiability. If v is differentiable, then the weak and the classical derivatives coincide. There are, however, functions that are not differentiable in the classical sense, but possess a weak derivative.

EXAMPLE 1.18. The absolute value function $v(x) = |x|$ on $\Omega = (-1, 1)$ is not differentiable on $(-1, 1)$. Yet, its weak derivative is given by

$$v'(x) = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x \geq 0. \end{cases} \tag{2}$$

Note that we can modify elements of $L^1_{\text{loc}}(\Omega)$ at $x = 0$ to any value.

From the example we see that functions with certain kinks can be weakly differentiable.

EXAMPLE 1.19. We subdivide the interval $(-1, 1)$ into finitely many sub-intervals $[x_j, x_{j+1}]$, where

$$-1 = x_1 < \cdots < x_N = 1 \quad \text{and} \quad \cup_{j=1}^{N-1} [x_j, x_{j+1}] = \bar{\Omega} = [-1, 1],$$

and consider the globally continuous functions that are affine when restricted to any of the sub-intervals $[x_j, x_{j+1}]$. Any such function is weakly differentiable.

The functions from the foregoing example allow for a very simple representation, and so they are generally suited for numerical computations. It is easy to verify that any such function can be characterized by the vector $(v(x_j))_{j=1}^N$ of its values at the points x_j. Between these nodal points, the values are interpolated by straight lines.
It is possible to generalize this construction to higher space dimensions. We only consider the case $n = 2$ in this lecture in order to minimize the technical efforts. Let the domain $\bar{\Omega}$ be subdivided in triangles. We consider the space of functions that are globally continuous and that are affine when restricted to any of the triangles. In order to define such spaces, we introduce a suitable class of triangular partitions.

DEFINITION 1.20 (triangle). A subset $T \subseteq \mathbb{R}^2$ is called *triangle* if there exists $(z_1, z_2, z_3) \in (\mathbb{R}^2)^3$ such that T is the convex hull of z_1, z_2, z_3 and these three points do not belong on one straight line. The points z_1, z_2, z_3 are called *vertices*. The line segments between z_j, z_k for $j \neq k$ are called *edges*. ♦

DEFINITION 1.21 (regular triangulation). Let $\mathcal{T} \subset 2^{\bar{\Omega}}$ be a finite set of triangles in $\bar{\Omega}$ ($2^{\bar{\Omega}}$ denotes the power set). The set $\mathcal{T}$ is called a *regular triangulation* of Ω if die the triangles cover the domain $\bar{\Omega}$, i.e., $\bigcup_{T \in \mathcal{T}} = \bar{\Omega}$, and if any pair $(T, K) \in \mathcal{T}^2$ satisfies one of the following relations:

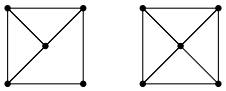
(i) $T \cap K = \emptyset$

(ii) $T \cap K$ is a common vertex
(iii) $T \cap K$ is a common edge
(iv) $T = K$.

♦

This means that the elements of a regular triangulation may only meet under certain rules.

EXAMPLE 1.22. A non-regular and a regular triangulation of the square:



In what follows, $\mathcal{T}$ will always denote a regular triangulation of Ω. Let $T \in \mathcal{T}$ be a triangle. The affine functions over T are denoted by

$$P_1(T) := \{v \in L^\infty(T) : \exists (a,b,c) \in \mathbb{R}^3 \forall x \in T,\ v(x) = a + bx_1 + cx_2\}.$$

The functions that are piecewise affine with respect to $\mathcal{T}$ (but possibly globally discontinuous) are denoted by

$$P_1(\mathcal{T}) := \{v \in L^\infty(\Omega) : \forall T \in \mathcal{T}, v|_T \in P_1(T)\}.$$

Finally, the continuous and piecewise affine functions are denoted by

$$S^1(\mathcal{T}) := C^0(\Omega) \cap P_1(\mathcal{T})$$

and the subspace with zero boundary conditions reads

$$S^1_0(\mathcal{T}) := \{v \in S^1(\mathcal{T}) : v|_{\partial\Omega} = 0\}.$$

The letter S shall remind us of *splines*; a notion that is possibly known from one-dimensional interpolation.
The following property is very important, and its proof is discussed in the problems below.

LEMMA 1.23. *The elements of $S^1(\mathcal{T})$ are weakly differentiable.*

PROOF. Exercise. □

The set of vertices (or *nodes*) of a triangle is denoted by $\mathcal{N}(T)$ and the set of all vertices is

$$\mathcal{N} := \{z \in \bar{\Omega} : \text{there exists } T \in \mathcal{T} \text{ having } z \text{ as a vertex}\} = \bigcup_{T \in \mathcal{T}} \mathcal{N}(T).$$

The basis we choose for $S^1(\mathcal{T})$ or $S^1_0(\mathcal{T})$ is the *nodal basis*. First, we define the nodal basis of $S^1(\mathcal{T})$ (no boundary conditions) by $(\varphi_z)_{z \in \mathcal{N}}$, where for any $z \in \mathcal{N}$ the function $\varphi_z \in S^1(\mathcal{T})$ is defined by the property

$$\varphi_z(y) = \delta_{yz} = \begin{cases} 1 & \text{if } y = z \\ 0 & \text{if } y \in \mathcal{N} \setminus \{z\}. \end{cases} \tag{3}$$

These functions are usually referred to as "hat functions". It will be shown in the exercises that these function indeed form a basis.
In order to define a nodal basis of $S^1_0(\mathcal{T})$, one omits the hat functions belonging to boundary vertices. To this end, we define the boundary vertices by $\mathcal{N}(\partial\Omega) := \partial\Omega \cap \mathcal{N}$ and the inner vertices by $\mathcal{N}(\Omega) := \mathcal{N} \setminus \mathcal{N}(\partial\Omega)$. The nodal basis of $S^1_0(\mathcal{T})$ then reads

$$(\varphi_z : z \in \mathcal{N}(\Omega)).$$

As in classical Lagrange interpolation, the coefficients with respect to the nodal basis are given by the nodal values. This means that any function $v_h \in S^1(\mathcal{T})$ can be expanded as follows

$$v_h = \sum_{z \in \mathcal{N}} v_h(z)\varphi_z.$$

The spaces $S^1(\mathcal{T})$ and $S^1_0(\mathcal{T})$ are called *finite element spaces*. Any continuous function $v \in C(\bar{\Omega})$ can be approximated by its interpolation $Iv \in S^1(\mathcal{T})$ as follows

$$Iv := \sum_{z \in \mathcal{N}} v(z)\varphi_z.$$

The map $I : C(\bar{\Omega}) \to S^1(\mathcal{T})$ is called *interpolation operator*. For the case of zero boundary conditions, the definition is analogous.
Let us now briefly discuss how to operate with triangulations and finite element functions on a computer (using Python).
We describe a triangulation by prescribing a list of nodes and a list of triangles. The nodes are put in a list `coord` $\in \mathbb{R}^{N\times 2}$. The x and y coordinate of the jth node are written to the jth row. In the example of Figure 1 this means

```
coord = np.asarray([[0,0],
                    [1,0],
                    [1,1],
                    [0,1],
                    [.5,.5]])
```

for the unit square $(0,1)^2$. Here, we use the library `numpy`:

```
import numpy as np
```

Now we form triangles out of the node numbers. We use convention that the numbering is counterclockwise. The list `triangles` $\in \mathbb{R}^{N\times 3}$ contains in its jth row the three node numbers of triangle number j. In the example from Figure 1 this reads

```
triangles = np.asarray([[0,1,4],
                        [1,2,4],
                        [2,3,4],
                        [3,0,4]])
```

We finally save the node pairs of the boundary edges on the Dirichlet boundary

```
dirichlet= np.array([[0,1],
                     [1,2],
                     [2,3],
                     [3,0]])
```

We will comment on (and make use of) this later. In Python we can now plot our triangulation by:

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.tri as mtri

plt.triplot(mtri.Triangulation(coord[:,0], coord[:, 1], triangles))
plt.show()
```

If we want to generate a surface plot of a piecewise affine function from $S^1(\mathcal{T})$, we can use `trisurf`. Figure 2 shows a complete example.
The triangulation in the above example is very coarse. Finer triangulations can be obtained by mesh refinement. A very simple refinement rule is called *red refinement*. Here, every triangle is subdivided in four congruent sub-triangles by connecting the edge midpoints by straight lines. We provide a routine `red_refine.py` on the

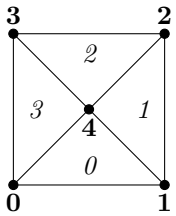


FIGURE 1. Triangulation of the square $(0,1)^2$ in four triangles. The bold numbers indicate the node numbers wile the numbers of the triangles are displayed in italic.

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.tri as mtri
from mpl_toolkits import mplot3d
from mpl_toolkits.mplot3d import Axes3D

coord = np.asarray([[0,0],[1,0],[1,1],[0,1],[.5,.5]])
triangles = np.asarray([[0,1,4],[1,2,4],[2,3,4],[3,0,4]])
dirichlet= np.array([[0,1],[1,2],[2,3],[3,0]])
# show triangulation
plt.triplot(mtri.Triangulation(coord[:,0], coord[:, 1], triangles))
plt.show()
# plot the interpolation of the function x+y
func = lambda x, y:  x + y
func2=np.vectorize(func)
z=func2(coord[:,0],coord[:,1])
fig = plt.figure(figsize =(14, 9))
ax = plt.axes(projection ='3d')
trisurf = ax.plot_trisurf(coord[:,0],coord[:,1],z,
                          triangles = triangles,
                          cmap =plt.get_cmap('summer'),
                          edgecolor='Gray');
plt.show()
```

FIGURE 2. Sample use of `triplot` and `trisurf`

lecture webpage. We do not care about the actual code, but we just use it. It can be used as follows

```
neumann=np.zeros([0, 2])
coord, triangles, dirichlet,_,_,_ = \
           red_refine(coord, triangles, dirichlet, neumann)
```

Here, `neumann` is just an empty list that, at this stage, has no importance. Later in the lecture we will also consider problems with a second type of boundary condition (so-called Neumann boundary condition), but for the moment we can ignore it; we also do not care about the three ignored output arguments of the function.

1.2.3. The finite element method. So far we did not study any existence theory to the Dirichlet problem of Poisson's equation. Instead we first introduce the central numerical method of this lecture: the finite element method (FEM). We want to use it to approximately solve Poisson's and other equations. In the course of this lecture we will then justify the method by convergence theory. But for the moment we motivate the scheme in a purely heuristic manner in order to be able to quickly proceed with practical results.

The point of departure for the FEM is that the energy functional J from the Dirichlet principle (Theorem 1.15) is well defined for elements from $S^1(\mathcal{T})$ or $S^1_0(\mathcal{T})$. Indeed, when considering the functional

$$J(v) = \frac{1}{2}\int_\Omega |\nabla v|^2 dx - \int_\Omega f v\, dx$$

we see that, for $v \in S^1(\mathcal{T})$, the gradient ∇v is defined in the sense of weak derivatives. It belongs to $L^2(\Omega)$ (it is even piecewise constant). Thus, the first part of the sum is finite. The second term is finite as well if we impose the (fairly weak) condition $f \in L^2(\Omega)$: the Hölder (or Cauchy-Schwarz) inequality then reveals

$$\left|\int_\Omega f v\, dx\right| \le \|f\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}.$$

Since we are considering the homogeneous Dirichlet problem (i.e., a zero boundary condition), we restrict the attention to approximations from the subspace $S^1_0(\mathcal{T})$. In our notation, we indicate that we are dealing with "discrete functions" by adding the index h to the variables. As an approximation to the solution u to the Dirichlet problem with right-hand side $f \in L^2(\Omega)$ we seek $u_h \in S^1_0(\mathcal{T})$ minimizing the functional J over the finite-dimensional space $S^1_0(\mathcal{T})$, written

$$u_h \in \operatorname*{arg\,min}_{v_h \in S^1_0(\mathcal{T})} J(v_h). \tag{4}$$

THEOREM 1.24. *Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded, polygonal Lipschitz domain with a regular triangulation $\mathcal{T}$. Given any $f \in L^2(\Omega)$, there exists a unique $u_h \in S^1_0(\mathcal{T})$ solving the discrete optimization problem* (4)*. Furthermore,* (4) *is equivalent to the condition*

$$\int_\Omega \nabla u_h \cdot \nabla v_h dx = \int_\Omega f v_h dx \quad \textit{for all } v_h \in S^1_0(\mathcal{T}). \tag{5}$$

PROOF. Like in the proof of Theorem 1.15 one can show that the minimization problem (4) leads to the (in this case discrete) Euler–Lagrange equation (5) as a necessary condition. Details will be worked out in the problem sessions. Let us now have a closer look at (5). Note carefully that the left-hand side of (5) is a positive definite bilinear form, while the right-hand side is a linear form (that is, an element of the dual space). This is shown as an exercise. It is actually immediate to see that the left-hand side is bilinear and positive semidefinite. The definiteness follows from the Dirichlet boundary condition. For a better overview, we now formulate (5) in terms of vectors and matrices.

Let the dimension of the finite-dimensional space $S^1_0(\mathcal{T})$ be denoted by $N \in \mathbb{N}$. Then $S^1_0(\mathcal{T})$ has the nodal basis $(\varphi_1, \dots \varphi_N) \in (S^1_0(\mathcal{T}))^N$. With respect to this basis, we can represent the parts

$$a(u_h, v_h) := \int_\Omega \nabla u_h \cdot \nabla v_h dx \quad \text{und} \quad F(v_h) := \int_\Omega f v_h dx$$

from (5) in the usual way as matrices as follows: Let $v_h = \sum_{j=1}^N x_j \varphi_j$ and $w_h = \sum_{k=1}^N y_k \varphi_k$ be elements of $S^1_0(\mathcal{T})$. We then have

$$a(v_h, w_h) = x^\top A y \quad \text{for } A \in \mathbb{R}^{N\times N} \text{ where } A_{jk} = a(\varphi_j, \varphi_k) \quad (j,k = 1, \dots, N).$$

Similarly, we have for $F \in V_h^*$ that

$$F(w_h) = b^\top y \quad \text{for } b \in \mathbb{R}^N \text{ where } b_k = F(\varphi_k) \quad (k = 1, \dots, N).$$

Here, $x, y \in \mathbb{R}^n$ are the vectors with entries x_j, y_k. If we expand $u_h = \sum_j x_j \varphi_j$ with respect to the given basis, then the coefficients x of the solution u_h (if existent)

satisfy the system

$$A^\top x = b. \tag{6}$$

This means that we have transformed (5) in the equivalent matrix-vector system (6). For this system it is immediate that it is uniquely solvable because A is positive definite (because $a(\cdot,\cdot)$ is). Thus, there exists a unique solution $u_h \in S^1_0(\mathcal{T})$ to (5). What is left to be shown is that u_h minimizes the functional J. Since A is symmetric, we derive with (6) the relation

$$\begin{aligned} J(u_h + w_h^y) - J(u_h) &= \frac{1}{2}(x+y)^\top A^\top (x+y) - (x+y)^\top b \\ &= \frac{1}{2} x^\top A^\top x - x^\top b + y^\top A^\top y \\ &= J(u_h) + y^\top A^\top y \geq J(u_h) \quad \text{for each } y \in \mathbb{R}^N \end{aligned}$$

where we write $w_h^y = \sum_{k=1}^N y_k \varphi_k$. The last estimate follows with the positive definiteness of A. Hence, J attains a minimum at u_h. The minimum is unique as the necessary condition (5) is satisfied by exactly one minimizer (namely u_h). □

In summary, we have formulated the finite-dimensional problem (4) as the linear system (6). Let us remark that the form (6) is more general in that it does not require symmetry but only definiteness. Let us now describe how to implement the FEM on the computer. For this, we need to assemble the matrix A and the vector b from the foregoing proof within a computer program.
In order to discretize Poisson's equation (subject to homogeneous Dirichlet boundary conditions) with the FEM, we need

- a triangulation $\mathcal{T}$, described through the data structures `coord`, `triangles`, `dirichlet`,
- the right-hand side f, e.g. given through values at certain points or as function,
- a vector b representing the linear functional $\int_\Omega f \bullet \, dx$ with respect to the nodal basis of $S^1(\mathcal{T})$,
- the so-called *stiffness matrix* A, i.e., the matrix representing the bilinear form from Poisson's equation with respect to the nodal basis of $S^1(\mathcal{T})$.

With these objects at hand, we can solve (6). It is important to restrict the matrices to the *degrees of freedom*. In our case, these correspond to the inner nodes (as the values for the boundary nodes are already fixed by the value 0). The list of degrees of freedom is usually given the variable name `dof`.
We start by specifying all required packages, see Figure 3. The structure of the program is displayed in Figure 4.
It remains to describe the routines for assembling the stiffness matrix A and the right-hand side vector b. We start with A. First, we build up *local* stiffness matrices for each triangle T

$$A_T^{loc} := \Big(\int_T \nabla\varphi_j \cdot \nabla\varphi_k \, dx\Big)_{j,k=1,2,3}.$$

Here, the vertices of T are locally numbered by $1, 2, 3$. Since the φ_j are affine functions, their gradients are constant so that we arrive at the formula

$$A_T^{loc} = \operatorname{area}(T) \begin{bmatrix} \nabla\varphi_1^\top \\ \nabla\varphi_2^\top \\ \nabla\varphi_3^\top \end{bmatrix} \begin{bmatrix} \nabla\varphi_1 & \nabla\varphi_2 & \nabla\varphi_3 \end{bmatrix}.$$

```
import numpy as np
from mpl_toolkits import mplot3d
import matplotlib.tri as mtri
import matplotlib.pyplot as plt
from mpl_toolkits import mplot3d
from mpl_toolkits.mplot3d import Axes3D
from mpl_toolkits.mplot3d import proj3d
import math
import pylab
from red_refine import red_refine #our refinement routine
import scipy.sparse
import scipy.sparse.linalg
from scipy.sparse import csr_matrix
```

FIGURE 3. The required packages in Python

```
def FEM(coord,triangles,dirichlet,f):
    nnodes=np.size(coord,0)
    A=stiffness_matrix(coord,triangles)
    b=RHS_vector(coord,triangles,f)
    dbnodes=np.unique(dirichlet)
    dof=np.setdiff1d(range(0,nnodes),dbnodes)
    A_inner=A[np.ix_(dof,dof)]
    b_inner=b[dof]
    x=np.zeros(nnodes)
    x[dof]=scipy.sparse.linalg.spsolve(A_inner,b_inner)
    return x
```

FIGURE 4. The basic FEM routine.

The area is easily computed as follows. With the three vertices $z_1, z_2, z_3 \in \mathbb{R}^2$ of T, we have that

$$\text{area}(T) = \frac{1}{2}\det[z_2 - z_1, z_3 - z_1].$$

For the computation of $\nabla\varphi_j$ we observe that the basis functions (or barycentric coordinates) satisfy the system

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \end{bmatrix}}_{\in\mathbb{R}^{3\times 3}} \begin{bmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \varphi_3(x) \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ x \end{bmatrix}}_{\in\mathbb{R}^{3\times 1}}$$

for any T. If we take derivatives (w.r.t. x) on both sides, we arrive at

$$\underbrace{\begin{bmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \end{bmatrix}}_{\in\mathbb{R}^{3\times 3}} \underbrace{\begin{bmatrix} \nabla\varphi_1^\top \\ \nabla\varphi_2^\top \\ \nabla\varphi_3^\top \end{bmatrix}}_{\in\mathbb{R}^{3\times 2}} = \underbrace{\begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}}_{\in\mathbb{R}^{3\times 2}}.$$

Therefore

$$\begin{bmatrix} \nabla\varphi_1^\top \\ \nabla\varphi_2^\top \\ \nabla\varphi_3^\top \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ z_1 & z_2 & z_3 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

We compute all local stiffness matrices in a loop

```
    nelems=np.size(triangles,0)
    Alocal=np.zeros((nelems,3,3))

    for j in range(0,nelems):
```

```
nodes_loc=triangles[j,:]
coord_loc=coord[nodes_loc,:]
T=np.array([coord_loc[1,:]-coord_loc[0,:] ,
       coord_loc[2,:]-coord_loc[0,:] ])
area = 0.5 * ( T[0,0]*T[1,1] - T[0,1]*T[1,0] )
T= np.concatenate((np.array([[1,1,1]]), coord_loc.T),axis=0)
T1= np.array([[0,0],[1,0],[0,1]])
grads = np.linalg.solve(T,T1)
Alocal[j,:,:]=area* np.matmul(grads,grads.T)
```

Now we need to assemble the local stiffness matrices into the global stiffness matrix. The entry A_{jk} of the global stiffness matrix is given by

$$A_{jk} = \int_\Omega \nabla\varphi_j \cdot \nabla\varphi_k \, dx = \sum_{T\in\mathcal{T}} \int_T \nabla\varphi_j \cdot \nabla\varphi_k \, dx = \sum_{\substack{T\in\mathcal{T} \\ \text{nodes } j,k \\ \text{belong to } K}} \int_T \nabla\varphi_j \cdot \nabla\varphi_k \, dx.$$

This means that, for any triangle, we save the index pairs (j, k in the above sum) assigning the global node numbers to the entries of the local stiffness matrix. We write these indices into the arrays `I1`, `I2`. We then build up a sparse matrix based on these indices (note that repeated indices imply summation).

```
nelems=np.size(triangles,0)
nnodes=np.size(coord,0)
I1=np.zeros((nelems,3,3))
I2=np.zeros((nelems,3,3))

for j in range(0,nelems):
    nodes_loc=triangles[j,:]
    I1[j,:,:] = np.concatenate((np.array([nodes_loc]),\
          np.array([nodes_loc]),np.array([nodes_loc])),axis=0)
    I2[j,:,:] = np.concatenate((np.array([nodes_loc]).T,\
            np.array([nodes_loc]).T,np.array([nodes_loc]).T),axis
                =1)

Alocal=np.reshape(Alocal,(9*nelems,1)).T
I1=np.reshape(I1,(9*nelems,1)).T
I2=np.reshape(I2,(9*nelems,1)).T
A=csr_matrix((Alocal[0,:],(I1[0,:],I2[0,:])),shape = (nnodes,
    nnodes))
```

The full routine for the stiffness matrix can be found in Figure 5. We now proceed with the assembling of the right-hand side. We again run a loop over all elements. Since b is not sparse, we can just update the vector in each loop iteration. For approximating the integral, we use the midpoint rule

$$\int_T f\varphi_j \, dx \approx \text{area}(T) f(m)\varphi_j(m),$$

where $m = \frac{1}{3}(z_1 + z_2 + z_3)$ is the midpoint (barycentre) of T. Since φ_j is affine, we can easily compute $\varphi_j(m) = 1/3$. This results in the routine of Figure 6.
For testing the FEM code, we use the above data for the unit square. In the code they will be loaded by a function `geom_square`. We use the following right-hand side for validation

$$f(x) = 2(x_1(1-x_1) + x_2(1-x_2)).$$

The exact solution reads

$$u(x) = x_1(x_1-1)x_2(x_2-1).$$

For a very basic convergence test for the L^∞ norm we now execute the following lines of code

```
def stiffness_matrix(coord,triangles):
    nelems=np.size(triangles,0)
    nnodes=np.size(coord,0)
    Alocal=np.zeros((nelems,3,3))
    I1=np.zeros((nelems,3,3))
    I2=np.zeros((nelems,3,3))

    for j in range(0,nelems):
        nodes_loc=triangles[j,:]
        coord_loc=coord[nodes_loc,:]
        T=np.array([coord_loc[1,:]-coord_loc[0,:] ,
               coord_loc[2,:]-coord_loc[0,:] ])
        area = 0.5 * ( T[0,0]*T[1,1] - T[0,1]*T[1,0] )
        tmp1= np.concatenate((np.array([[1,1,1]]), coord_loc.T),axis
            =0)
        tmp2= np.array([[0,0],[1,0],[0,1]])
        grads = np.linalg.solve(tmp1,tmp2)
        Alocal[j,:,:]=area* np.matmul(grads,grads.T)
        I1[j,:,:] = np.concatenate((np.array([nodes_loc]),np.array([
            nodes_loc]),np.array([nodes_loc])),axis=0)
        I2[j,:,:] = np.concatenate((np.array([nodes_loc]).T,np.array([
            nodes_loc]).T,np.array([nodes_loc]).T),axis=1)

    Alocal=np.reshape(Alocal,(9*nelems,1)).T
    I1=np.reshape(I1,(9*nelems,1)).T
    I2=np.reshape(I2,(9*nelems,1)).T
    A=csr_matrix((Alocal[0,:],(I1[0,:],I2[0,:])),shape = (nnodes,
        nnodes))
    return A
```

FIGURE 5. Routine for the stiffness matrix.

```
def RHS_vector(coord,triangles,f):
    nelems=np.size(triangles,0)
    nnodes=np.size(coord,0)
    b=np.zeros(nnodes)
    for j in range(0,nelems):
        nodes_loc=triangles[j,:]
        coord_loc=coord[nodes_loc,:]
        tmp=np.array([coord_loc[1,:]-coord_loc[0,:] ,
               coord_loc[2,:]-coord_loc[0,:] ])
        area = 0.5 * ( tmp[0,0]*tmp[1,1] - tmp[0,1]*tmp[1,0] )
        mid=1/3*(coord_loc[0,:]+coord_loc[1,:]+coord_loc[2,:])
        b[nodes_loc]=b[nodes_loc]+area/3*f(mid[0],mid[1])
    return b
```

FIGURE 6. Routine for the right-hand side vector.

```
fun = lambda x, y:  (x-x**2)*(y- y**2)
u_exact=np.vectorize(fun)
f = lambda x, y:  2* ((x-x**2)+(y- y**2) )
coord, triangles, dirichlet, neumann = get_geom()
max_err=np.zeros(5)
for j in range(0,5):
    coord, triangles, dirichlet,_,_,_ = \
           red_refine(coord, triangles, dirichlet, neumann)
    x=FEM(coord, triangles, dirichlet,f)
    u_at_nodes=u_exact(coord[:,0],coord[:,1])
    max_err[j]=np.max(np.abs(u_at_nodes-x))
```

```
print(max_err)
```

1.2.4. Elementary properties of Sobolev spaces. We introduce spaces of functions that posses appropriate weak derivatives. It will turn out that these are suited for a sound theory of Poisson's equation (and similar problems). We shall prove many, but not all of the stated results.

DEFINITION 1.25 (Sobolev spaces). Let $\Omega \subseteq \mathbb{R}^2$ be bounded and open. Define

$$H^1(\Omega) := \{v \in L^2(\Omega) : \forall j \in \{1,2\}\ \partial_j v \in L^2(\Omega)\}.$$

That is, the functions from $H^1(\Omega)$ belong to $L^2(\Omega)$; their first weak derivatives exist and belong to $L^2(\Omega)$ as well. ♦

REMARK 1.26. For L^p spaces instead of L^2 one can analogously define Sobolev spaces, which are commonly denoted by $W^{1,p}(\Omega)$. We will not consider such spaces in this lecture. ♦

REMARK 1.27. In many cases it will be enough for our purposes to confine ourselves to polygonal Lipschitz domains. Most of the results will, however, hold under weaker conditions. ♦

Sobolev functions have far more structure than generic L^2 functions. Recall that elements from $L^2(\Omega)$ are equivalence classes (up to equality almost everywhere) and that point evaluations are not well defined. This is generally the case for Sobolev function, too. Yet, we will see that such functions possess boundary values in some generalized sense. We first study an important property, namely that $H^1(\Omega)$ can equivalently be defined by a completion process. Let us define the following norm on $H^1(\Omega)$,

$$\|v\|_{H^1(\Omega)} := \sqrt{\|v\|^2_{L^2(\Omega)} + \|\nabla v\|^2_{L^2(\Omega)}}.$$

REMARK 1.28. We use the convention that $\|\nabla v\|^2_{L^2(\Omega)} = \int_\Omega |\nabla v|^2\, dx$ for the Euclidean norm $|\cdot|$. ♦

THEOREM 1.29. *Let $\Omega \subseteq \mathbb{R}^2$ be open and bounded. The space $H^1(\Omega)$ is complete with respect to the norm $\|\cdot\|_{H^1(\Omega)}$, i.e. a Banach space.*

PROOF. The proof is left as an exercise (Problem 1.28). □

In the following, we will use arguments involving the open covering

$$U_j := \{x \in \Omega : \frac{\operatorname{diam}(\Omega)}{2} 2^{-j} \le \operatorname{dist}(x, \partial\Omega) \le 2 \operatorname{diam}(\Omega) 2^{-j}\}$$

for the open and bounded set Ω. This covering is locally finite in the sense that any of the sets U_j has a nonempty intersection with only finitely many sets U_k (exercise). It is known from multivariate analysis that there exists a corresponding smooth partition of unity, that is a family $(\eta_j)_j$ of nonnegative functions $\eta_j \in C_c^\infty(U_j)$ with the property

$$\sum_j \eta_j(x) = 1 \quad \text{for all } x \in \Omega.$$

THEOREM 1.30 (approximation by smooth functions I). *Let $\Omega \subseteq \mathbb{R}^2$ be open and bounded. Then the space $H^1(\Omega)$ is the completion of*

$$C^\infty(\Omega) \cap H^1(\Omega)$$

with respect to the norm $\|\cdot\|_{H^1(\Omega)}$. In other words: Given any $v \in H^1(\Omega)$ there exists a sequence $(v_j)_j$ in $C^\infty(\Omega) \cap H^1(\Omega)$ with the property that $\|v - v_j\|_{H^1(\Omega)} \to 0$ for $j \to \infty$.

PROOF. Let $v \in H^1(\Omega)$ and $\varepsilon > 0$. The proof uses approximation by convolution, which is known from integration theory. Let $(\psi_\varepsilon)_{\varepsilon>0}$ be a standard Dirac sequence. For $v \in H^1(\Omega)$ we define the approximation

$$v_\varepsilon(x) := (\psi_\varepsilon * 1_\Omega v)(x) = \int_\Omega \psi_\varepsilon(x-y)v(y)dy,$$

where "dy" means integration w.r.t. the Lebesgue measure and the variable y. Given any open subset $D \subseteq \Omega$ with $\delta := \operatorname{dist}(D, \partial\Omega) > 0$ we then have

$$v_\varepsilon \in H^1(D) \cap C^\infty(D) \quad \text{provided } \varepsilon < \delta.$$

Let us prove this claim. It is known that $v_\varepsilon \in C^\infty(D)$ as well as $v_\varepsilon \in L^2(D)$. Moreover, the partial derivatives satisfy due to rotational symmetry of ψ_ε and Lebesgue's theorem

$$\partial_j v_\varepsilon(x) = \int_\Omega \frac{\partial}{\partial x_j}\psi_\varepsilon(x-y)v(y)dy = -\int_\Omega \frac{\partial}{\partial y_j}\psi_\varepsilon(x-y)v(y)dy.$$

We observe that, for any $x \in \Omega$ with $\operatorname{dist}(x, \partial\Omega) > \varepsilon$, the function $y \mapsto \psi_\varepsilon(x-y)$ belongs to $C_c^\infty(\Omega)$. By definition of the weak derivative ∂_j we thus have for such x, after integration by parts, that,

$$\partial_j v_\varepsilon(x) = \int_\Omega \psi_\varepsilon(x-y)\partial_j v(y)dy = (\psi_\varepsilon * 1_\Omega \partial_j v)(x).$$

In other words: The derivative of the regularization is the regularized derivative. By known results from measure and integration theory related to approximation by convolution we thus infer convergence $v_\varepsilon \to v$ in $L^2(D)$ and $\partial_j v_\varepsilon \to \partial_j v$ in $L^2(D)$, and therefore $v_\varepsilon \to v$ in $H^1(D)$ for $\varepsilon \to 0$.
In order to show the result on Ω (and not just on the subsets D) another technical step is required. Let $(U_k)_{k\in\mathbb{N}}$ be a locally finite open covering of Ω and $(\eta_k)_{k\in\mathbb{N}}$ be a corresponding smooth partition of unity as constructed above. Owing to the above results we can find, for any k and any $\varepsilon > 0$, an approximation $v_{k,\varepsilon} \in C^\infty(U_k)$ by convolution such that

$$\|v - v_{k,\varepsilon}\|_{H^1(U_k)} \le \frac{1}{10}\frac{\varepsilon}{2^k(1 + \|\eta_k\|_{C^1(\bar\Omega)})}.$$

We combine these local approximations and define

$$v_{\Omega,\varepsilon} := \sum_{k\in\mathbb{N}_0} \eta_k v_{k,\varepsilon}.$$

We compute with the product rule (ses also Problem 1.29)

$$\partial_j(v - v_{\Omega,\varepsilon}) = \sum_{k\in\mathbb{N}_0} \partial_j(\eta_k(v - v_{k,\varepsilon})) = \sum_{k\in\mathbb{N}_0} (\partial_j\eta_k(v - v_{k,\varepsilon}) + \eta_k\partial_j(v - v_{k,\varepsilon}))$$

and obtain with the triangle inequality

$$\begin{aligned}\|\partial_j(v - v_{\Omega,\varepsilon})\|_{L^2(\Omega)} &\le \sum_{k\in\mathbb{N}_0} (\|\partial_j\eta_k(v - v_{k,\varepsilon})\|_{L^2(U_k)} + \|\eta_k\partial_j(v - v_{k,\varepsilon})\|_{L^2(U_k)}) \\ &\le 2\sum_{k\in\mathbb{N}_0} \|\eta_k\|_{C^1(\bar\Omega)}\|v - v_{k,\varepsilon}\|_{H^1(U_k)}\end{aligned}$$

Here we have estimated the terms containing η_k by their maxima; we furthermore used the elementary estimate $a+b \le 2\sqrt{a^2+b^2}$ for real a, b. With the above choice of $v_{k,\varepsilon}$ and the geometric series we arrive at

$$\|\partial_j(v - v_{\Omega,\varepsilon})\|_{L^2(\Omega)} \le \frac{2}{5}\varepsilon.$$

For the L^2 norm we obtain in a similar fashion the direct estimate

$$\|v - v_{\Omega,\varepsilon}\|_{L^2(\Omega)} \leq \sum_{k\in\mathbb{N}_0} \|\eta_k\|_{C^1(\bar\Omega)} \|v - v_{k,\varepsilon}\|_{L^2(U_k)} \leq \varepsilon/5.$$

Altogether

$$\|v - v_{\Omega,\varepsilon}\|_{H^1(\Omega)} \leq \|v - v_{\Omega,\varepsilon}\|_{L^2(\Omega)} + \sum_{j=1}^{2} \|\partial_j (v - v_{\Omega,\varepsilon})\|_{L^2(\Omega)} \leq \frac{\varepsilon}{5} + \frac{2\varepsilon}{5} + \frac{2\varepsilon}{5} \leq \varepsilon.$$

We have shown that, given any $v \in H^1(\Omega)$, there exists an approximation $v_{\Omega,\varepsilon}$ that converges in the $H^1(\Omega)$-Norm towards v as $\varepsilon \to 0$. □

THEOREM 1.31 (approximation by smooth functions II). *Let $\Omega \subseteq \mathbb{R}^2$ be an open and bounded Lipschitz domain. Then, $H^1(\Omega)$ is the completion of*

$$C^\infty(\bar\Omega)$$

with respect to the norm $\|\cdot\|_{H^1(\Omega)}$. In other words: Given any $v \in H^1(\Omega)$ there exists a sequence $(v_j)_j$ in $C^\infty(\bar\Omega)$ with the property that $\|v - v_j\|_{H^1(\Omega)} \to 0$ for $j \to \infty$.

PROOF. In contrast to Theorem 1.30 we need some regularity of the boundary. Since the domain has a Lipschitz boundary, there are open sets $U^1, \dots, U^N$ covering a neighbourhood U of $\partial\Omega$ and having the property (after some shift and rotation of the coordinate system) that

$$\Omega \cap U^j \subset \{x \in U^j : x_2 > \gamma(x_1)\} \tag{7}$$

as well as $\partial\Omega \cap U^j = \mathrm{Graph}(\gamma)$ for some Lipschitz function γ. We choose smooth functions $\phi_j \in C_c^\infty(U^j)$, that form a partition of unity on U (i.e., $\sum_{j=1}^N \phi_j = 1$ in U). In order to cover the inner part of Ω, we choose an open set $U^0 \subset\subset \Omega$ such that $\Omega \subseteq \cup_{j=0}^N U_j$ and set $\phi_0 := 1 - \sum_{j=1}^N \phi_j$. It follows that there is an open domain $\hat\Omega \supset\supset \Omega$ for which $(\phi_j : j = 0, \dots, N)$ is a partition of unity. Now we claim that for any $v \in H^1(\Omega)$, any $\varepsilon > 0$, and any $j = 0, \dots, N$ there exists some $w_j \in C_c^\infty(\mathbb{R}^2)$ such that $\|\phi_j v - w_j\|_{H^1(U_j)} \leq \varepsilon/(N+1)$. Assuming for the moment this property, we define $w := \sum_{j=0}^N w_j$. We then immediately infer $w \in C_c^\infty(\mathbb{R}^2)$ and, in particular, $w|_\Omega \in C^\infty(\bar\Omega)$. The triangle inequality furthermore implies

$$\|v - w\|_{H^1(\Omega)} \leq \sum_{j=0}^{N} \|\phi_j v - w_j\|_{H^1(\Omega)} \leq \varepsilon$$

which proves the assertion of the theorem. Let us now prove the above claim. For $j = 0$ the claim follows as in Theorem 1.30. Given any $j \in \{1, \dots, N\}$, we choose a local coordinate system according to (7). We extend $v_j := \phi_j v$ by zero to the whole $\mathbb{R}^2$. For small $t > 0$ we then consider the shifted function $v_{j,t}(x) = v_j(x + t\begin{pmatrix}0\\1\end{pmatrix})$, whose support locally overlaps beyond $\partial\Omega$. As in the proof of Theorem 1.30 we can approximate $v_{j,t}$ via convolution by functions $\psi_\eta * v_{j,t}$ (these satisfy $(\psi_\eta * v_{j,t})|_{U^j} \in C^\infty(\bar U_j)$). In particular, we have $\|v_{j,t} - \psi_\eta * u_{j,t}\|_{H^1(\Omega)} \to 0$ for $\eta \to 0$. On the other hand, we have $\|v_{j,t} - v_j\|_{H^1(\Omega)} \to 0$ for $t \to 0$ (see Problem 1.30). For any prescribed $\delta > 0$ we thus conclude

$$\|v_{j,t} - v_j\|_{H^1(\Omega)} < \delta/2 \quad \text{for sufficiently small } t > 0.$$

Next, we choose $\eta > 0$ small enough such that

$$\|v_{j,t} - \psi_\eta * v_{j,t}\|_{H^1(\Omega)} < \delta/2$$

and obtain with the triangle inequality that

$$\|v_j - \psi_\eta * v_{j,t}\|_{H^1(\Omega)} < \delta.$$

□

1.2.5. Traces; the Dirichlet problem in Sobolev spaces. In the previous lecture we have seen hat suitable smooth functions are dense in $H^1(\Omega)$. As a first application of this result we will show that we can assign boundary values to functions from $H^1(\Omega)$ in a consistent fashion. Such property is, obviously, impossible to achieve for mere $L^2(\Omega)$ functions.

THEOREM 1.32 (trace identity and trace inequality for triangles). *Let $T \subseteq \mathbb{R}^2$ be a triangle with some edge $F \subseteq T$ and opposite vertex $P \in T$. Any function $v \in C^1(T)$ then satisfies*

$$\frac{|T|}{|F|}\int_F v\,ds = \int_T v\,dx + \frac{1}{2}\int_T (\bullet - P)\cdot\nabla v\,dx$$

and

$$\|v\|^2_{L^2(F)} \le \frac{3|F|}{2|T|}\|v\|^2_{L^2(T)} + \frac{|F|}{2|T|}\operatorname{diam}(T)^2\|\nabla v\|^2_{L^2(T)}.$$

Here, $|T|$ denotes the area of T and $|F|$ denotes the length of F.

PROOF. We have $\operatorname{div}(\bullet - P) = 2$ (in two space dimensions). Integration by parts therefore reveals

$$\int_T v\,dx + \frac{1}{2}\int_T (\bullet - P)\cdot\nabla v\,dx = \int_{\partial T} v\,(\bullet - P)\cdot\nu\,ds,$$

where ν is the outer unit normal of T. We observe that, on the two edges of T different from F, the vector $(\bullet - P)$ is tangential to ∂T and, thus, its product with ν equals zero. Hence,

$$\int_{\partial T} v\,(\bullet - P)\cdot\nu\,ds = \int_F v\,(\bullet - P)\cdot\nu\,ds.$$

Since furthermore ν is constant along F, the quantity $(\bullet - P)\cdot\nu$ is constant on F as well, and its value corresponds to the orthogonal projection of $(\bullet - P)$ in direction of ν. This is precisely the length of the height on F, which by elementary geometry takes the value $2|T|/|F|$. This proves the first assertion.
In order to show the second claimed property, we apply the trace identity to v^2. Note that $\nabla(v^2) = 2v\nabla v$. We thus infer

$$\frac{|T|}{|F|}\int_F v^2\,ds = \int_T v^2\,dx + \int_T (\bullet - P)\cdot v\nabla v\,dx \le \int_T v^2\,dx + \operatorname{diam}(T)\int_T |v|\,|\nabla v|\,dx,$$

where in the second step we have estimated the length of $(\bullet - P)$ by the diameter of T. After rearranging the identity we obtain

$$\|v\|^2_{L^2(F)} \le \frac{|F|}{|T|}\|v\|^2_{L^2(T)} + \frac{|F|}{|T|}\operatorname{diam}(T)\int_T |v|\,|\nabla v|\,dx.$$

We use the Cauchy-Schwarz inequality and the Young inequality $2ab \le a^2 + b^2$ to estimate the second integral as follows

$$\begin{aligned}\operatorname{diam}(T)\frac{|F|}{|T|}\int_T |v|\,|\nabla v|\,dx &= \frac{|F|}{|T|}\int_T |v|\,\big(\operatorname{diam}(T)|\nabla v|\big)\,dx \\ &\le \frac{|F|}{2|T|}(\|v\|^2_{L^2(T)} + \operatorname{diam}(T)^2\|\nabla v\|^2_{L^2(T)}).\end{aligned}$$

This implies the second assertion. □

THEOREM 1.33. *Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded domain with polygonal Lipschitz boundary. Then, there exists a unique continuous and linear map $S : H^1(\Omega) \to L^2(\partial\Omega)$ with the property*

$$Sv = v|_{\partial\Omega} \quad \text{for all } v \in H^1(\Omega) \cap C^0(\bar{\Omega}).$$

REMARK 1.34. A linear map $T : H^1(\Omega) \to L^2(\partial\Omega)$ is said to be continuous if there exists a constant $C_T < \infty$ such that

$$\|Tv\|_{L^2(\partial\Omega)} \leq C_T \|v\|_{H^1(\Omega)} \quad \text{for all } v \in H^1(\Omega).$$

The theorem states the following. The operation of taking boundary values, which is well defined for functions from $C^0(\bar{\Omega})$, has a unique continuation to functions from $H^1(\Omega)$. Taking such generalized boundary values still leads to functions in $L^2(\partial\Omega)$, and we interpret these as boundary values of functions from $H^1(\Omega)$. This concept turns out important if we wish to pose the Dirichlet problem in Sobolev spaces. The operator S is called *trace operator*, and Sv is called the trace of v on $\partial\Omega$. ♦

PROOF OF THEOREM 1.33. We tesselate $\bar{\Omega}$ with a regular triangulation $\mathcal{T}$. Given any $v \in C^1(\bar{\Omega})$, we can apply Theorem 1.32 to any boundary edge $F \subseteq \partial\Omega$ of the triangulation and obtain

$$\|v\|^2_{L^2(F)} \leq \frac{3|F|}{2|T_F|}\|v\|^2_{L^2(T_F)} + \frac{|F|}{2|T_F|}\operatorname{diam}(T_F)^2\|\nabla v\|^2_{L^2(T_F)}.$$

Here, T_F is the uniquely defined triangle containing F as an edge. Since $\mathcal{T}$ has finitely many elements, the constant

$$C_{\mathcal{T}} := \max\left\{\max\left\{\frac{3|F|}{2|T_F|}, \frac{|F|}{2|T_F|}\operatorname{diam}(T_F)^2\right\} : F \subseteq \partial\Omega \text{ edge with triangle } T_F\right\}$$

is finite and we have the estimate

$$\|v\|^2_{L^2(F)} \leq C_{\mathcal{T}}(\|v\|^2_{L^2(T_F)} + \|\nabla v\|^2_{L^2(T_F)}) \quad \text{for all } v \in C^1(\bar{\Omega}).$$

For the whole boundary $\partial\Omega$ we then obtain

$$\|v\|^2_{L^2(\partial\Omega)} = \sum_{\substack{F \text{ boundary edge} \\ \text{of } \mathcal{T}}} \|v\|^2_{L^2(F)} \leq C_{\mathcal{T}} \sum_{\substack{F \text{ boundary edge} \\ \text{of } \mathcal{T}}} (\|v\|^2_{L^2(T_F)} + \|\nabla v\|^2_{L^2(T_F)}).$$

Obviously, every triangle can contain (at most) three boundary edges. Thus, any T_F occurs at most three times in the sum on the right hand side, and we can estimate

$$\|v\|^2_{L^2(\partial\Omega)} \leq 3C_{\mathcal{T}} \sum_{T\in\mathcal{T}} (\|v\|^2_{L^2(T)} + \|\nabla v\|^2_{L^2(T)}) = 3C_{\mathcal{T}}\|v\|^2_{H^1(\Omega)}.$$

Altogether, we have shown that there is a constant C such that

$$\|v\|_{L^2(\partial\Omega)} \leq C\|v\|_{H^1(\Omega)} \quad \text{for all } v \in C^1(\bar{\Omega}).$$

Thus, we have shown the desired estimate for the map $S : C^1(\bar{\Omega}) \to L^2(\Omega)$ assigning boundary values on the space $C^1(\bar{\Omega})$, which is dense in $H^1(\Omega)$ by Theorem 1.31. Thus, by an elementary result of linear functional analysis, there is a unique continuation of S to $H^1(\Omega)$. The continuity constant remains the same. □

As a consequence from the trace theorem, it makes sense to impose boundary values on functions from $H^1(\Omega)$. We will usually write $u|_{\partial\Omega}$ instead of Su etc., but we need to be aware that this function is only of class L^2 on $\partial\Omega$. For the Dirichlet problem, it is reasonable to consider the following subspace

$$H^1_0(\Omega) := \{v \in H^1(\Omega) : v|_{\partial\Omega} = 0\},$$

i.e., the space of Sobolev functions with zero boundary values. Sometimes, an alternative characterization turns out useful.

THEOREM 1.35. *Let $\Omega \subseteq \mathbb{R}^2$ be an open and bounded Lipschitz domain. Then $H_0^1(\Omega)$ is the closure of $C_c^\infty(\Omega)$ with respect to the norm $\|\cdot\|_{H^1(\Omega)}$, i.e.,*

$$H_0^1(\Omega) := \overline{C_c^\infty(\Omega)}^{\|\cdot\|_{H^1(\Omega)}}.$$

PROOF. We do not prove this technical result here. Its proof relies on similar techniques as Theorem 1.30 or Theorem 1.31 and can be found in the literature [**Dob10, Eva10**]. □

For functions from $H_0^1(\Omega)$, the L^2 norm can be controlled by the L^2 norm of the gradient. This result is called Friedrichs' inequality (sometimes Poincaré–Friedrichs inequality).

THEOREM 1.36 (Friedrichs' inequality). *Let Ω be an open, bounded, and connected Lipschitz domain. Then there exists a constant $C > 0$ such that*

$$\|v\|_{L^2(\Omega)} \leq C\|\nabla v\|_{L^2(\Omega)} \quad \text{for all } v \in H_0^1(\Omega).$$

The constant is C proportional to the diameter of Ω.

PROOF. The proof is left as an exercise. We sketch the basic idea. In view of Theorem 1.35, it is enough to consider $v \in C_c^\infty(\Omega)$ and then argue by density. We extend v by zero to some larger rectangular box containing Ω. After shifting coordinates, we may assume that $\Omega \subseteq (0, L)^2$, $L > 0$. Then, v is of class $C_c^\infty((0, L)^2)$ with respect to this box. For any $x \in \Omega$, we can integrate

$$v(x) = v(x_1, x_2) = v(0, x_2) + \int_0^{x_1} \partial_1 v(t, x_2)\, dt.$$

We observe that the boundary term is zero. For the remaining term, we use the Cauchy-Schwarz/Hölder inequality and obtain

$$|v(x)|^2 \leq L \int_0^L |\partial_1 v(t, x_2)|^2\, dt.$$

We now intergrate with respect to x_1

$$\int_0^L |v(x)|^2 dx_1 \leq L^2 \int_0^L |\partial_1 v(t, x_2)|^2\, dt.$$

and thereafter integrate with respect to x_2

$$\int_0^L \int_0^L |v(x)|^2\, dx_1 dx_2 \leq L^2 \int_0^L \int_0^L |\partial_1 v(t, x_2)|^2\, dt dx_2.$$

Since the support of v lies inside Ω, this implies the asserted estimate for v. By a density argument, it is true for all functions from $H_0^1(\Omega)$. □

The most important implication of Friedrichs' inequality is that $\|\nabla \cdot\|_{L^2(\Omega)}$ defines a norm on $H_0^1(\Omega)$. (Convince yourself that this cannot be a norm on the larger space $H^1(\Omega)$ by considering constant functions.) Denoting the constant from Friedrichs' inequality by C_F, we indeed have the equivalence of norms

$$\|v\|_{H^1(\Omega)}^2 = \|v\|_{L^2(\Omega)}^2 + \|\nabla v\|_{L^2(\Omega)}^2 \leq (1 + C_\mathrm{F}^2)\|\nabla v\|_{L^2(\Omega)}^2 \leq (1 + C_\mathrm{F}^2)\|v\|_{H^1(\Omega)}^2. \tag{8}$$

We use the notation $|v|_1 = \|\nabla v\|_{L^2(\Omega)}$.
We are now in the position to formulate the Dirichlet problem in Sobolev spaces. It is based on the necessary condition from Theorem 1.15 (Dirichlet principle).

DEFINITION 1.37. Let $\Omega \subseteq \mathbb{R}^2$ be an open and bounded Lipschitz domain. Given $f \in L^2(\Omega)$, the variational (or weak) formulation of the Dirichlet problem for Poisson's equation seeks $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx \quad \text{for all } v \in H_0^1(\Omega).$$

♦

This generalizes Poisson's equation in the sense that every classical solution will also be a solution to the variational formulation (see exercises).
We will now establish that the variational formulation possesses unique solutions. This will be an immediate consequence of the Riesz representation theorem, a fundamental result from the theory of Hilbert spaces. This topic is taught in any class on linear functional analysis, and we briefly recall basic details.
Let X be a (real) linear space. Given a symmetric and positive definite bilinear form $(\cdot,\cdot)_X$, we define $\|x\|_X = \sqrt{(x,x)_X}$ for any $x \in X$. It is elementary to establish the Cauchy–Schwarz inequality

$$(x,y)_X \leq \|x\|_X \, \|y\|_X \quad \text{for any } x, y \in X.$$

It can be shown that $\|\cdot\|_X$ defines a norm on X (thereby justifying the notation).

DEFINITION 1.38 (Hilbert space). A linear space X (over $\mathbb{R}$) equipped with a symmertic and positive definite bilinear form $(\cdot,\cdot)_X$ is called *Hilbert space* if it is complete with respect to the norm $\|\cdot\|_X := \sqrt{(\cdot,\cdot)_X}$. ♦

Basically, Hilbert spaces are Banach spaces with an euclidean structure. We recall the dual space X^*, the space of continuous linear functionals over X. The Reisz representation theorem states that there exists an isometric isomorphism between X and X^*.

THEOREM 1.39 (Riesz representation theorem). *Let X be a Hilbert space with inner product $(\cdot,\cdot)_X$ and let $F \in X^*$ be a continuous linear functional. Then there exists a unique element $x \in X$ with the property*

$$(x,y)_X = F(y) \quad \textit{for all } y \in X.$$

The element x satisfies $\|x\|_X = \|F\|_{X^*}$.

PROOF. The proof is taught in every course on linear functional analysis. □

We now use Hilbert space methods to show well-posedness of our variational formulation.

LEMMA 1.40. *Let $\Omega \subseteq \mathbb{R}^2$ be an open and bounded Lipschitz domain. The space $H_0^1(\Omega)$ equipped with the bilinear form*

$$\int_\Omega \nabla v \cdot \nabla w \, dx$$

is a Hilbert space.

PROOF. Friedrichs' inequality shows that the symmetric bilinear form is positive definite. The completeness with respect to $|\cdot|_1$ is a consequence of the equivalence of norms (8) and the fact that $H_0^1(\Omega)$ is a closed subspace of $H^1(\Omega)$. □

THEOREM 1.41. *Let $\Omega \subseteq \mathbb{R}^2$ be an open and bounded Lipschitz domain and let $f \in L^2(\Omega)$. The variational formulation of the Dirichlet problem of Poisson's equation has a unique solution $u \in H_0^1(\Omega)$.*

PROOF. We check that

$$v \mapsto \int_\Omega f v \, dx$$

is a continuous linear functional on the Hilbert space $H_0^1(\Omega)$. This follows from the Cauchy and the Friedrichs inequality

$$\int_\Omega f v \, dx \le \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} \le \|f\|_{L^2(\Omega)} C_{\mathrm{F}} |v|_1.$$

Hence, we are in the setting of the Riesz representation theorem, which states that there is a unique element $u \in H_0^1(\Omega)$ satisfying

$$\int_\Omega \nabla u \cdot \nabla v \, dx = \int_\Omega f v \, dx \quad \text{for all } v \in H_0^1(\Omega).$$

□

By using elementary Hilbert space theory we could establish existence and uniqueness to the Dirichlet problem for any right-hand side $f \in L^2(\Omega)$. Note that this setting only needs the weak form of the Laplacian; furthermore even situations like Example 1.13 are covered by the theory. What is furthermore attractive about this approach is a direct characterization of the finite element error. We note that $S_0^1(\mathcal{T})$ is a finite-dimensional subspace of the Hilbert space $H_0^1(\Omega)$. It turns out that the finite element solution $u_h \in S_0^1(\mathcal{T})$ is the orthogonal projection of u to $S_0^1(\mathcal{T})$ and, thus, the best approximation in this space.

THEOREM 1.42. *Let $\Omega \subset \mathbb{R}^2$ be an open, bounded, connected Lipschitz polygon with a triangulation $\mathcal{T}$. Given $f \in L^2(\Omega)$, the error between the solution $u \in H_0^1(\Omega)$ to the variational form of Poisson's equation and the finite element solution $u_h \in S_0^1(\mathcal{T})$ satisfies*

$$|u - u_h|_1 = \inf_{v_h \in S_0^1(\mathcal{T})} |u - v_h|_1.$$

PROOF. We observe that, for any $v_h \in S_0^1(\mathcal{T}) \subseteq H_0^1(\Omega)$, we have

$$\begin{aligned}\int_\Omega \nabla(u - u_h) \cdot \nabla v_h \, dx &= \int_\Omega \nabla u \cdot \nabla v_h \, dx - \int_\Omega \nabla u_h \cdot \nabla v_h \, dx \\ &= \int_\Omega f v_h \, dx - \int_\Omega f v_h \, dx = 0.\end{aligned}$$

This property is also called *Galerkin orthogonality* because it describes that the error is orthogonal on discrete space. We compute

$$|u - u_h|_1^2 = \int_\Omega \nabla(u - u_h) \cdot \nabla u \, dx - \int_\Omega \nabla(u - u_h) \cdot \nabla u_h \, dx.$$

and see from the Galerkin orthogonality that the second term equals zero and remains zero if u_h on the right is replaced by any $v_h \in S_0^1(\mathcal{T})$. Thus,

$$|u - u_h|_1^2 = \int_\Omega \nabla(u - u_h) \cdot \nabla(u - v_h) \, dx \le |u - u_h|_1 |u - v_h|_1$$

by Cauchy's inequality. The assertion then follows from dividing by $|u - u_h|_1$ and taking the infimum over v_h. □

We have seen that the finite element method is, in some sense, optimal. The result should illustrate the basic idea of the error analysis. In the next sections, we will generalize the theory to more general operators (not just the Laplacian) and see that the finite element method satisfies similar error bounds. It will turn out as a special case of *Galerkin approximations*.

1.2.6. Finite element theory for linear coercive operators. We can use Hilbert space methods to consider more complicated second-order operators than the Laplacian. In many applications, we encounter PDEs of the form

$$-\operatorname{div}(A\nabla u) + b\cdot\nabla u + cu = f$$

for a matrix field A, a vector field b, and a function c. These three terms are referred to as *diffusion*, *advection*, and *reaction*, respectively. As for the Laplacian, we can interpret the divergence operator weakly and derive the following variational formulation for $u \in H^1_0(\Omega)$:

$$\int_\Omega \Big((A\nabla u)\cdot\nabla v + (b\cdot\nabla u)v + c\,uv \Big)\,dx = \int_\Omega fv\,dx \quad \text{for all } v \in H^1_0(\Omega). \tag{9}$$

In this section we will study under which (sufficient) conditions this system has a unique solution. Note that the left-hand side need not be symmetric, and an immediate use of scalar products like in the case of Poisson's equation is not possible. Note furthermore that there need not be any related energy functional or Dirichlet principle.
The following important result extends, in some sense, the Riesz representation theorem to a class of nonsymmetric bilinear forms.

THEOREM 1.43 (Lax–Milgram lemma). *Let V be a real Hilbert space with inner product $(\cdot,\cdot)_V$ and let $a : V\times V \to \mathbb{R}$ be a bilinear form satisfying the following two properties*

- $\exists\beta > 0\,\forall (v,w)\in V^2 \quad |a(v,w)| \le \beta\|v\|_V\|w\|_V$ *(continuity)*
- $\exists\alpha > 0\,\forall v\in V \quad \alpha\|v\|_V^2 \le a(v,v)$ *(coercivity)* .

Then, there exists a unique map $T : V \to V$ with the property

$$a(w,v) = (Tw,v)_V \quad \text{for all } (v,w)\in V^2.$$

The map T is linear, continuous, and invertible with

$$\|T\|_{L(V,V)} \le \beta \quad \text{and} \quad \|T^{-1}\|_{L(V,V)} \le \frac{1}{\alpha}.$$

PROOF. We will prove a more general result later in this class. It will imply the Lax–Milgram lemma. □

COROLLARY 1.44. *Let a be a continuous and coervice bilinear form on a Hilbert space V with inner product $(\cdot,\cdot)_V$. Given any $F\in V^*$, there is a unique $u\in V$ such that*

$$a(u,v) = F(v) \quad \text{for all } v\in V.$$

It satisfies $\|u\|_V \le \alpha^{-1}\|F\|_{L(V,V)}$.

PROOF. Let $f\in V$ denote the den Riesz representative of F in V, and let T denote the mapping from the Lax–Milgram lemma. Then, $u := T^{-1}f$ satisfies

$$F(v) = (f,v)_V = (TT^{-1}f,v)_V = (Tu,v)_V = a(u,v)$$

for any $v\in V$. The norm bound for u folllows from the bound on T^{-1} from the Lax–Milgram lemma. □

EXAMPLE 1.45 (general elliptic operator). Let

$$A\in[L^\infty(\Omega)]^{2\times 2}, \quad b\in[L^\infty(\Omega)]^2, \quad c\in L^\infty(\Omega)$$

be the coefficients of the above PDE with $f\in L^2(\Omega)$ und homogeneous Dirichlet boundary condition. After multipying with test functions an integrating (by parts) we obtain the following weak formulation: Seek $u\in H^1_0(\Omega)$ such that

$$a(u,v) = F(v) \quad \text{for all } v\in H^1_0(\Omega),$$

where

$$a(u,v) := \int_\Omega \Big((A\nabla u)\cdot\nabla v + (b\cdot\nabla u)v + c\,uv \Big)\,dx \quad \text{and} \quad F(v) := \int_\Omega f v\,dx.$$

We now apply, under further structural assumptions, the Lax–Milgram lemma to the above setting.

THEOREM 1.46. *Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded, connected Lipschitz polygon. Let the coefficients A, b, c from Example 1.45 satisfy the following assumptions.*

- *The field A is pointwise symmetric and there exist real numbers $0 < a_0, a_1$ such that*
$$a_0|\xi|^2 \le (A(x)\xi)\cdot\xi \le a_1|\xi|^2 \qquad \text{a.e. in } \Omega \text{ for all } \xi \in \mathbb{R}^2$$
i.e., A is uniformly positive definite.
- *The vector field b is divergence-free, $\operatorname{div} b = 0$ (in the sense of the weak divergence, see Problem 1.25).*
- *The function $c \ge 0$ is nonnegative.*

Then there exists a unique solution $u \in H^1_0(\Omega)$ to the weak form from Example 1.45. It satisfies the bound
$$|u|_1 \le C\|f\|_{L^2(\Omega)}$$
for some constant $C > 0$ that is independent of f and u.

PROOF. The proof is left as an exercise. It is enough to verify that a satisfies the assumptions from the Lax–Milgram lemma. □

A finite element discretization of this problem is straight-forward: We restrict the bilinear form a and the right-hand side to finite element functions. The form a, however, need not be a scalar product, and the best-approximation property (as in the Laplacian case) is not valid in its original form. We will now study approximations in a more general setting.

DEFINITION 1.47. Let a be a coercive and continuous bilinear form on a Hilbert space V, and let $V_h \subseteq V$ be a closed subspace. Given $F \in V^*$, let $u \in V$ solve
$$a(u,v) = F(v) \quad \text{for all } v \in V.$$
The unique solution $u_h \in V_h$ to
$$a(u_h, v_h) = F(v_h) \quad \text{for all } v_h \in V_h$$
is called the *Galerkin approximation* to u. ⧫

REMARK 1.48. We remark that, in the foregoing definition, the Galerkin approximation indeed exists and is unique. This follows from the fact that closed subspaces of Hilbert spaces are again Hilbert spaces. It is immediate to see that the Lax–Milgram lemma applies on such subspaces as well. ⧫

EXAMPLE 1.49. The finite element approximation to Poisson's equation is a Galerkin method based on the finite-dimensional subspace $V_h := S^1_0(\mathcal{T})$ of $V := H^1_0(\Omega)$. The finite element approximation to the operator from Example 1.45 is a Galerkin method as well.

We now formulate the basic error estimate for Galerkin approximations.

THEOREM 1.50 (Céa's lemma). *Let V be a Hilbert space and let $V_h \subseteq V$ be a closed subspace. Let $a : V \times V \to \mathbb{R}$ be a continuous and coervice bilinear form (with α, β as in the Lax–Milgram lemma) and let $F \in V^*$. Let $u \in V$ solve*
$$a(u,v) = F(v) \quad \textit{for all } v \in V.$$

The Galerkin approximation $u_h \in V_h$ solving

$$a(u_h, v_h) = F(v_h) \quad \textit{for all } v_h \in V_h$$

satisfies the following error bound

$$\|u - u_h\|_V \leq \frac{\beta}{\alpha} \inf_{v_h \in V_h} \|u - v_h\|_V.$$

PROOF. The coercivity reveals the following relation of the norm and the form a,

$$\alpha \|u - u_h\|_V^2 = a(u - u_h, u - u_h). \tag{10}$$

We observe that, due to the property $V_h \subseteq V$, the variational problems in V and V_h satisfy

$$a(u - u_h, v_h) = a(u, v_h) - a(u_h, v_h) = F(v_h) - F(v_h) = 0. \tag{11}$$

As a consequence, we can take an arbitrary $w_h \in V_h$ and compute

$$\begin{aligned} a(u - u_h, u - u_h) &= a(u - u_h, u) - \underbrace{a(u - u_h, u_h)}_{=0} = a(u - u_h, u) \\ &= a(u - u_h, u) - \underbrace{a(u - u_h, w_h)}_{=0} = a(u - u_h, u - w_h). \end{aligned}$$

We use this relation in the above formula (10) and deduce from the continuity that

$$\alpha \|u - u_h\|_V^2 = a(u - u_h, u - w_h) \leq \beta \|u - u_h\|_V \|u - w_h\|_V.$$

In case that $u - u_h = 0$, the assertion of the theorem is trivially satisfied. Otherwise, we can divide by the norm of $u - u_h$ and take the infimum over w_h. □

Due to the factor β/α in the error estimate, the Galerkin method is said to be quasi-optimal. We can apply the abstract setting to the finite element method for the second-order system from Example 1.45 and Theorem 1.46 and obtain the quasi-optimal bound

$$|u - u_h|_1 \leq C \inf_{v_h \in S_0^1(\mathcal{T})} |u - v_h|_1.$$

In this sense, the finite element method computes a near-best approximation in the space $S_0^1(\mathcal{T})$. We will quantify this approximation in the forthcoming sections.
Let us now discuss in which regards the theory and methods to more general situations.

Inhomogeneous Dirichlet values. It is often required to prescribe nonzero boundary values to the Dirichlet problem. For some given function $u_D : \partial\Omega \to \mathbb{R}$ one is interested in finding a function u satisfying

$$-\Delta u = f \text{ in } \Omega \quad \text{and} \quad u = u_D \text{ on } \partial\Omega.$$

In the variational form one seeks $u \in H^1(\Omega)$ with

$$a(u, v) = \int_\Omega f v \, dx \text{ for all } v \in H_0^1(\Omega) \quad \text{and} \quad u = u_D \text{ on } \partial\Omega \text{ in the sense of traces.}$$

Here, a is the Laplacian inner product, but the generalization to other operators is immediate. Of course, for this formulation to make sense, the data u_D must belong to the range of the trace operator, i.e., it must possess a continuation to a function from $H^1(\Omega)$. We denote the range of the trace operator by $H^{1/2}(\partial\Omega)$, without further characterizing it here. Let $\hat{u}_D \in H^1(\Omega)$ denote an extension of u_D to the domain Ω. The idea is to shift the solution by u_D and to seek for $w = u - u_D$, which then has zero boundary data. One solves for $w \in H_0^1(\Omega)$ such that

$$a(w, v) = \int_\Omega f v \, dx - a(u_D, v) \quad \text{for all } v \in H_0^1(\Omega).$$

It follows from the continuity of a that the right-hand side defines a continuous linear functional on $H^1_0(\Omega)$ and so there is a unique solution w. Then, $u := w + u_D$ solves the inhomogeneous boundary value problem. In a practical finite element implementation, we proceed analogously. It might be required to interpolate u_D with piecewise affine functions along the boundary so that it is the trace of a finite element function. We can extend this finite element function to the domain by setting it to zero at all interior vertices and denote the coefficient vector by x_D. The modified right-hand side then reads $\tilde{b} := b - A^\top x_D$, where b is the load vector related to f and A is the stiffness matrix. We then solve for x_0 (which is zero at the boundary vertices) by restricting the system $A^\top x_0 = \tilde{b}$ to the interior vertices as the degrees of freedom. Then, $x := x_0 + x_D$ is the coefficient vector of the finite element solution.

Neumann boundary values. In many applications, for example when u describes the heat distribution in some domain Ω, one wants to prescribe $(A\nabla u) \cdot \nu$ on the boundary rather than actual values for u. In the context of a heat distribution, this corresponds to the heat flux. The boundary is then subdivided in two disjoint parts

$$\partial\Omega = \Gamma_D \cup \Gamma_N$$

where Γ_D is relatively closed. The part Γ_D is called the Dirichlet boundary and Γ_N is called the Neumann boundary. Either of the parts is allowed to be empty. The boundary value problem in its strong form then reads

$$-\operatorname{div} A\nabla u + b \cdot \nabla u + cu = f \text{ in } \Omega, \quad u = u_D \text{ on } \Gamma_D, \quad (A\nabla u) \cdot \nu = g \text{ on } \Gamma_N$$

where u_D is the prescribed Dirichlet data and $g \in L^2(\Gamma_N)$ is a given function, the so-called Neumann data. Assume for simplicity that $u_D = 0$. As we have no homogeneous boundary condition on the whole boundary, we need to work with the space

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

Be aware that in our integration-by-parts arguments the boundary terms do not vanish any more because test functions may be nonzero along Γ_N. Indeed, we find

$$\int_\Omega (-\operatorname{div} A\nabla u) v \, dx = \int_\Omega (A\nabla u) \cdot \nabla v \, dx - \int_{\Gamma_N} (A\nabla u) \cdot \nu \, v \, ds$$

for any test function $v \in H^1_D(\Omega)$. The term $(A\nabla u)\cdot\nu$ is prescribed by the Neumann data g. The weak formulation with Neumann data thus reads: Find $u \in H^1_D(\Omega)$ such that

$$a(u, v) = \int_\Omega fv \, dx + \int_{\Gamma_N} gv \, ds.$$

In our Python code, we can prescribe the Neumann boundary by the structure `neumann`, which so far had been left empty. In order to show well-posedness with the Lax–Milgram lemma, we need coercivity of a on $H^1_D(\Omega)$. This means that we need a generalization of Friedrichs' inequality to the case where our functions only vanish on some part (of positive surface measure) of the boundary. In the pure Neumann case $\Gamma_N = \partial\Omega$ and $\Gamma_D = \emptyset$, it is easy to see that there will be no unique solution because solutions may be shifted by arbitrary constants. In this case we therefore need to normalize the solution

$$\int_\Omega u \, dx = 0 \quad \text{in case of pure Neumann bounary conditions.}$$

This condition guarantees coercivity, as will be shown in the next theorem. We use the notation

$$H^1(\Omega)/\mathbb{R} = \{v \in H^1(\Omega) : \int_\Omega v \, dx = 0\}.$$

THEOREM 1.51 (generalized Poincaré and Friedrichs inequalities). *Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded, connected Lipschitz domain. There is a constant $C_P > 0$ such that*

$$\|v\|_{L^2(\Omega)} \leq C_P \|\nabla v\|_{L^2(\Omega)} \quad \text{for all } v \in H^1(\Omega)/\mathbb{R} \quad \text{(Poincaré inequality).}$$

Let $\Gamma_D \subseteq \partial\Omega$ have positive surface measure. Then there is a constant $C_F > 0$ such that

$$\|v\|_{L^2(\Omega)} \leq C_F \|\nabla v\|_{L^2(\Omega)} \quad \text{for all } v \in H^1_D(\Omega) \quad \text{(Friedrichs inequality).}$$

The constants C_F and C_P are proportional to the diameter of the domain Ω.

PROOF. The proof is based on a so-called compactness argument that will be presented in the next section. We postpone the proof to the exercises of that section. □

1.2.7. Finite element error estimates. We would like to quantify the right-hand side of Céa's lemma in terms of the mesh-size (maximum diameter of the triangles in $\mathcal{T}$). The idea is to plug in a suitable approximation in the infimum for which we then derive quantified bounds. To achieve this, we will use the finite element interpolation. It is, however, not a well defined on $H^1(\Omega)$ because it takes point evaluations, which need not exist without further assumptions (see Problem 1.8). This means that the interpolation operator, denoted by I_h, assigning the finite element interpolation $I_h v$ to any suitable (say continuous) function v, is not well defined on $H^1(\Omega)$, see Problem 1.41. We have seen in Problem 1.33 that point evaluations are well-defined in the space

$$H^2(\Omega) = \{v \in L^2(\Omega) : \text{all weak derivatives of } v \text{ up to order 2 exist in } L^2(\Omega)\}$$

with norm

$$\|v\|_{H^2(\Omega)} = \sqrt{\sum_{|\alpha| \leq 2} \|\partial^\alpha v\|^2_{L^2(\Omega)}}.$$

The proof, which was shown for triangles, can be extended to polygonal domains.

THEOREM 1.52. *Let $\Omega \subseteq \mathbb{R}^2$ be an open and bounded Lipschitz polygon. Then, we have the continuous embedding $H^2(\Omega) \hookrightarrow C(\bar{\Omega})$ and there exists a constant $C > 0$ such that*

$$\|v\|_{L^\infty(\Omega)} \leq C \|v\|_{H^2(\Omega)} \quad \text{for any } v \in H^2(\Omega).$$

PROOF. The proof is left as an exercise. The idea is to triangulate the domain and to use the bound that was proven for triangles. □

We have seen that we can apply the finite element interpolation I_h under the assumption that our solution u satisfies the stronger property $u \in H^1_0(\Omega) \cap H^2(\Omega)$. For a derivation of a quantitative bound on the interpolation under this assumption, we need an additional property from the theory of Sobolev spaces.

THEOREM 1.53. *Let $\Omega \subset \mathbb{R}^2$ be an open and bounded Lipschitz domain. Then, the embedding $H^1(\Omega) \hookrightarrow L^2(\Omega)$ is compact. That is, any weakly convergent sequence $v_n \rightharpoonup v$ $(n \to \infty)$ in $H^1(\Omega)$ converges strongly in $L^2(\Omega)$.*

PROOF. The proof is shown in advanced courses on linear functional analysis and is beyond the scope of this lecture. □

REMARK 1.54. An iterative application of Theorem 1.53 shows that $H^2(\Omega)$ is compactly embedded in $H^1(\Omega)$. ♦

With these tools we can now prove an interpolation error estimate on triangles. The constant appearing in the estimate will depend on the aspect ratio of the triangles.

DEFINITION 1.55. Let $T \subseteq \mathbb{R}^2$ be a triangle. Let h_T denote its diameter and let ρ_T denote the diameter of the largest ball inscribed to T. The quantity h_T/ρ_T is called the *aspect ratio* of T. ♦

The quantities ρ_T, h_T arise from the transformation of the domain. Observe that any pair of triangles T, $\hat{T}$ allows for some affine diffeomorphism $\Phi : \hat{T} \to T$, which can be written $\Phi(\hat{x}) = B\hat{x} + c$ with a 2×2 matrix $B = D\Phi$ and a vector c.

LEMMA 1.56. *Let $\Phi(\hat{x}) = B\hat{x} + c$ denote the affine map from a triangle $\hat{T}$ to the triangle T. Then, the spectral norm $\|\cdot\|$ of B and B^{-1} satisfies*

$$\|B\| \leq \frac{h_T}{\rho_{\hat{T}}} \quad \text{and} \quad \|B^{-1}\| \leq \frac{h_{\hat{T}}}{\rho_T}.$$

PROOF. Given any vector $\xi \in \mathbb{R}^2$ of length $|\xi| = \rho_{\hat{T}}$, there exists pair of points $\hat{x}$, $\hat{y}$ inside $\hat{T}$ with $\hat{x} - \hat{y} = \xi$ because the full ball of diameter $\rho_{\hat{T}}$ is contained in $\hat{T}$. Since $\Phi(\hat{x})$ and $\Phi(\hat{y})$ belong to T, the image under B satisfies $B\xi = B(\hat{x} - \hat{y}) = \Phi(\hat{x}) - \Phi(\hat{y})$ and its length is bounded by the diameter h_T. We thus compute

$$\|B\| = \sup_{\xi \in \mathbb{R}^2, |\xi|=1} |B\xi| = \sup_{\xi \in \mathbb{R}^2, |\xi| = \rho_{\hat{T}}} \frac{1}{\rho_{\hat{T}}} |B\xi| \leq \frac{h_T}{\rho_{\hat{T}}}.$$

The second asserted estimate follows from interchanging the roles of T and $\hat{T}$. □

We now prove the interpolation error estimate.

THEOREM 1.57 (interpolation error estimate). *There exists a constant $C > 0$ such that for any triangle $T \subseteq \mathbb{R}^2$ the interpolation error satisfies*

$$\|\nabla(v - I_h v)\|_{L^2(T)} \leq C \frac{h_T}{\rho_T} h_T \|D^2 v\|_{L^2(T)}$$
$$\text{and} \qquad \|v - I_h v\|_{L^2(T)} \leq C h_T^2 \|D^2 v\|_{L^2(T)}$$

for any $v \in H^2(T)$. Here, $\|D^2 v\|_{L^2(T)} = \sqrt{\int_T \sum_{j,k=1}^2 |\partial_{jk} v|^2 \, dx}$.

PROOF. We first prove an auxiliary estimate on some fixed reference triangle $\hat{T}$. We claim that there is a constant $\hat{C}$ such that any $w \in H^2(\hat{T})$ satisfies

$$\|w\|_{H^1(\hat{T})} \leq \hat{C}(\|D^2 w\|_{L^2(\hat{T})} + \sum_{z \in \mathcal{N}(\hat{T})} |w(z)|).$$

Assume for contradiction that the statement is wrong. Then there is a sequence $w_n \in H^2(\hat{T})$ with

$$\|w_n\|_{H^1(\hat{T})} \geq n(\|D^2 w_n\|_{L^2(\hat{T})} + \sum_{z \in \mathcal{N}(\hat{T})} |w_n(z)|) \quad \text{for all } n \in \mathbb{N}.$$

After normalizing the sequence to $\|w_n\|_{H^1(\hat{T})} = 1$ we obtain

$$\|D^2 w_n\|_{L^2(\hat{T})} + \sum_{z \in \mathcal{N}(\hat{T})} |w_n(z)| \leq 1/n \quad \text{for all } n \in \mathbb{N}.$$

The space $H^2(T)$ is reflexive, whence there exists a weakly convergent subsequence of this bounded sequence with some weak limit $w \in H^2(\hat{T})$. We do not relabel the subsequence and still denote it by w_n. The compact embedding of Theorem 1.53 shows that we have $w_n \to w$ in $H^1(\hat{T})$. It is even a Cauchy sequence in $H^2(\hat{T})$ because

$$\begin{aligned} \|w_j - w_k\|_{H^2(\hat{T})}^2 &= \|w_j - w_k\|_{H^1(\hat{T})}^2 + \|D^2(w_j - w_k)\|_{L^2(\hat{T})}^2 \\ &\leq \|w_j - w_k\|_{H^1(\hat{T})}^2 + \|D^2 w_j\|_{L^2(\hat{T})}^2 + \|D^2 w_k\|_{L^2(\hat{T})}^2 \end{aligned}$$

and the norms of the Hessian converge to 0. Therefore we have strong convergence $w_n \to w$ in $H^2(\hat{T})$, and $D^2 w = 0$. Thus, w is an affine function. By continuity, we furthermore see that $w(z) = 0$ at the vertices of $\hat{T}$. Thus, w is the zero function. But this contradicts $\|w_n\|_{H^1(\hat{T})} = 1$. This proves the claimed auxiliary estimate.
Now, let T be an arbitrary triangle. Then, there is an affine transformation

$$\Phi : \hat{T} \to T$$

from the reference triangle to T. We denote by $e := v - I_h v$ the interpolation error and observe from the change-of-variables formula that

$$\|\nabla e\|^2_{L^2(T)} = \int_T |\nabla e|^2 \, dx = \int_{\hat{T}} |(\nabla e) \cdot \Phi|^2 |\det D\Phi| \, dx$$

We use notation $\hat{e} := e \circ \Phi$. The chain rule reveals for any $\hat{x} \in \hat{T}$ that

$$\nabla \hat{e}(\hat{x}) = D\Phi(\hat{x})^\top \nabla e|_{\Phi(\hat{x})}.$$

Multiplying with the inverse of $D\Phi^\top$ and taking squares thus leads to

$$|(\nabla e) \circ \Phi|^2 = |D\Phi^{-\top} \nabla \hat{e}|^2 \leq \|D\Phi^{-1}\|^2 |\nabla \hat{e}|^2$$

where $\|\cdot\|$ denotes the (pointwise) spectral matrix norm.
We observe that $D\Phi$ is constant on $\hat{T}$ (because Φ is affine). We thus obtain

$$\|\nabla e\|^2_{L^2(T)} \leq \|D\Phi^{-1}\|^2 |\det D\Phi| \|\nabla \hat{e}\|^2_{L^2(\hat{T})}.$$

By the auxiliary result, there exists a constant $\hat{C}$, depending on $\hat{T}$, such that

$$\|\nabla \hat{e}\|^2_{L^2(\hat{T})} \leq \hat{C}^2 \|D^2 \hat{e}\|^2_{L^2(\hat{T})}.$$

Here, we have used that e, the interpolation error vanishes at the vertices of T, and so does the transformed function on the vertices of $\hat{T}$. So far we have shown

$$\|\nabla e\|^2_{L^2(T)} \leq \hat{C}^2 \|D\Phi^{-1}\|^2 \int_{\hat{T}} |D^2 \hat{v}|^2 |\det D\Phi| \, dx.$$

The chain rule shows

$$D^2 \hat{v}(\hat{x}) = D\Phi(\hat{x})^\top D^2 v|_{\Phi(\hat{x})} D\Phi(\hat{x}).$$

We thus find

$$|D^2 \hat{v}|^2 \leq \|D\Phi(\hat{x})\|^4 |(D^2 v) \circ \Phi|^2.$$

After transforming back to T we thus obtain

$$\|\nabla e\|^2_{L^2(T)} \leq \hat{C}^2 \|D\Phi^{-1}\|^2 \|D\Phi\|^4 \|D^2 \hat{v}\|^2_{L^2(T)}.$$

The norms of $D\Phi$ and its inverse can be estimated with Lemma 1.56 as follows

$$\|D\Phi^{-1}\|^2 \|D\Phi\|^4 \leq \frac{h_{\hat{T}}^2}{\rho_T^2} \frac{h_T^4}{\rho_{\hat{T}}^4} = \frac{h_{\hat{T}}^2}{\rho_{\hat{T}}^4} \frac{h_T^4}{\rho_T^2}.$$

The terms related to $\hat{T}$ are independent of T and can be estimated by some universal constant. We thus obtain (after taking squareroots) the asserted bound on the norm of the gradient. The bound on the L^2 norm is left as an exercise. □

We see from the interpolation error estimate of Theorem 1.57 that the interpolation error is proportional to h_T provided the aspect ratio of the triangle is bounded. If we take, for instance, any fixed triangle and refine it uniformly with the red refinement rule, the aspect ratio is bounded by a universal constant. We say that a family of triangulations with bounded aspect ratio is *shape-regular*. The approximation of an H^2 function is then determined by the mesh-size h_T and thus improved under mesh-refinement.

COROLLARY 1.58 (global interpolation error estimate). . *Let $\Omega \subseteq \mathbb{R}^2$ be an open and bounded polygonal Lipschitz domain. Let $\{\mathcal{T}_h\}_h$ be a shape-regular family of triangulations. Then, there is a constant $C > 0$ such that for any $v \in H^2(\Omega)$ the finite element interpolation I_h with respect to a mesh $\mathcal{T}_h$ satisfies*

$$\|\nabla(v - I_h v)\|_{L^2(\Omega)} \leq Ch\|D^2 v\|_{L^2(\Omega)} \quad \text{and} \quad \|v - I_h v\|_{L^2(\Omega)} \leq Ch^2\|D^2 v\|_{L^2(\Omega)}$$

for the maximal mesh-size $h = \max T \in \mathcal{T}_h h_T$.

PROOF. This follows from writing the L^2 norm as

$$\|\cdot\|_{L^2(\Omega)} = \sqrt{\sum_{T\in\mathcal{T}_h} \|\cdot\|^2_{L^2(T)}}$$

and using the local bounds of Theorem 1.57. □

We have seen that any $v \in H^2(\Omega)$ is approximated with order h by the finite element interpolation the H^1 norm and with order h^2 in the L^2 norm.
The combination with Céa's lemma now yields a quantified bound for the finite element approximation on our PDE.

COROLLARY 1.59. *Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded, connected Lipschitz polygon. Let the coefficients A, b, c satisfy the assumptions from Theorem 1.46. Assume that the solution $u \in H^1_0(\Omega)$ to the weak form from Example 1.45 additionally satisfies*

$$u \in H^1_0(\Omega) \cap H^2(\Omega).$$

Then, the error between u and the finite element approximation u_h with respect to a triangulation $\mathcal{T}_h$ from a shape-regular family satisfies

$$|u - u_h|_1 \leq Ch\|D^2 u\|_{L^2(\Omega)}.$$

PROOF. Céa's lemma (Theorem 1.50) yields

$$|u - u_h|_1 \leq \frac{\beta}{\alpha} \inf_{v_h \in \mathcal{S}^1_0(\mathcal{T}_h)} |u - v_h|_1$$

where α, β denote the coercivity and continuity constant, respectively. We now plug the choice $v_h := I_h u$ in the infimum. Note that the interpolation exists because $u \in H^2(\Omega)$ was assumed. The assertion then follows from the interpolation error estimate of Corollary 1.58. □

As an immediate question we ask under what condition the assumption $u \in H^2(\Omega)$ is satisfied. This is the topic of *regularity theory* and is beyond our discussion within this lecture. We only mention one important result here for the case of the Laplacian.

THEOREM 1.60 (regularity on convex domains). *Let $\Omega \subset \mathbb{R}^2$ be an open convex domain. Given any $f \in L^2(\Omega)$, the solution to the Dirichlet problem of the Laplacian (Poisson's equation) satisfies $u \in H^1_0(\Omega) \cap H^2(\Omega)$ with the bound*

$$\|D^2 u\|_{L^2(\Omega)} \leq C\|f\|_{L^2(\Omega)}.$$

PROOF. See the literature, e.g., [**Dob10, Eva10**]. □

When the domain is nonconvex, the solution may fail to belong to $H^2(\Omega)$. This is for instance the case in Example 1.13.
We can prove an improved bound for the error in the L^2 norm.

THEOREM 1.61 (L^2 error bound). *Let $\Omega \subset \mathbb{R}^2$ be an open convex domain. Given any $f \in L^2(\Omega)$, the solution to the Dirichlet problem of the Laplacian (Poisson's equation) and its finite element approximation satisfy*

$$\|u - u_h\|_{L^2(\Omega)} \leq Ch^2\|D^2 u\|_{L^2(\Omega)} \leq Ch^2\|f\|_{L^2(\Omega)}$$

PROOF. The technique employed in the proof is known as the *Aubin-Nitsche duality trick*. The idea is to solve for a solution $z \in H_0^1(\Omega)$ an auxiliary problem whose right-hand side is given by the error $e := u - u_h$. Let z solve

$$\int_\Omega \nabla z \cdot \nabla v \, dx = \int_\Omega ev \, dx \quad \text{for all } v \in H_0^1(\Omega).$$

We test the equation with $v := e$ and obtain

$$\|e\|_{L^2(\Omega)}^2 = \int_\Omega e\, e \, dx \quad \text{for all } v \in H_0^1(\Omega) = \int_\Omega \nabla e \cdot \nabla z \, dx.$$

We now use the Galerkin orthogonality and plug in the finite element approximation z_h to z,

$$\int_\Omega \nabla e \cdot \nabla z \, dx = \int_\Omega \nabla(u - u_h) \cdot \nabla z \, dx = \int_\Omega \nabla(u - u_h) \cdot \nabla(z - z_h) \, dx.$$

Corollary 1.59 implies for the finite element errors that

$$\begin{aligned} &\|\nabla(u - u_h)\|_{L^2(\Omega)} \le C\|D^2 u_h\|_{L^2(\Omega)} \\ \text{and} \qquad &\|\nabla(z - z_h)\|_{L^2(\Omega)} \le C\|D^2 z_h\|_{L^2(\Omega)} \le C\|e\|_{L^2(\Omega)}. \end{aligned}$$

We now combine the above formulas and divide by the norm of e to arrive at the first asserted estimate. The second one follows from Theorem 1.60. □

For simplicity, we have considered right-hand sides $f \in L^2(\Omega)$. The argument in Theorem 1.46 however even applies to any right-hand side F in the dual space of $H_0^1(\Omega)$. This dual space is denoted by

$$H^{-1}(\Omega) := [H_0^1(\Omega)]^*$$

where we use the norm $|\cdot|_1$. Accordingly, the norm in $H^{-1}(\Omega)$ reads

$$\|F\|_{H^{-1}(\Omega)} = \sup_{v \in H_0^1(\Omega)\setminus\{0\}} \frac{\langle F, v\rangle}{\|\nabla v\|_{L^2(\Omega)}}.$$

The space $L^2(\Omega)$ is embedded in $H^{-1}(\Omega)$ by the identification of $f \in L^2(\Omega)$ with the functional $T_f \in H^{-1}(\Omega)$ defined by

$$\langle T_f, v\rangle = \int_\Omega fv \, dx \quad \text{for all } v \in H_0^1(\Omega).$$

The map $f \mapsto T_f$ is injective and continuous, thus an embedding. The injectivity follows from the fact that

$$\int_\Omega gv \, dx \quad \text{for all } v \in H_0^1(\Omega)$$

implies $g = 0$ (by density of $H_0^1(\Omega)$ in $L^2(\Omega)$) whence T_g is zero in $H^{-1}(\Omega)$. Continuity follows from the Friedrichs inequality as follows:

$$\begin{aligned} \|T_f\|_{H^{-1}(\Omega)} = \sup_{v \in H_0^1(\Omega)\setminus\{0\}} \frac{\int_\Omega fv \, dx}{\|\nabla v\|_{L^2(\Omega)}} &\le \sup_{v \in H_0^1(\Omega)\setminus\{0\}} \frac{\|f\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}}{\|\nabla v\|_{L^2(\Omega)}} \\ &\le C_F\|f\|_{L^2(\Omega)}. \end{aligned}$$

This map thus leads to

$$H_0^1(\Omega) \subseteq L^2(\Omega) \hookrightarrow H^{-1}(\Omega)$$

and the embedding is often interpreted as the inclusion $L^2(\Omega) \subseteq H^{-1}(\Omega)$.

1.3. Quasi-interpolation

In contrast to the nodal interpolation from prior sections, the quasi-interpolation is well defined for any function from $H^1(\Omega)$. The idea is to replace point evaluations by certain volume averages around the vertices.
Given a triangulation $\mathcal{T}$ of the domain Ω, we define the nodal patch of any vertex $z \in \mathcal{N}$ by

$$\omega_z := \cup\{T \in T : z \in T\}$$

as the domain of all elements containing z, and the element patch

$$\omega_T := \cup_{z\in\mathcal{N}(T)}\omega_z$$

as the domain of all elements surrounding T. Given any $v \in H^1_0(\Omega)$, we then define its quasi-interpolation $R_h v \in S^1_0(\mathcal{T})$ by

$$R_h v := \sum_{z\in\mathcal{N}(\Omega)} \fint_{\omega_z} v\,dx\;\varphi_z$$

where we use the notation $\fint_{\omega_z}\cdot dx = |\omega_z|^{-1}\int_{\omega_z}\cdot dx$ for the integral mean. The operator $R_h : H^1_0(\Omega) \to S^1_0(\mathcal{T})$ is called *quasi-interpolation operator*. Its difference to the nodal interpolation is that the point evaluation $v(z)$ is replaced by the computation of some average around z.

THEOREM 1.62. *Let $\mathcal{T}$ be a regular triangulation of some open and bounded Lipschitz polygon Ω. There exists a constant $C > 0$ such that the quasi-interpolation R_h satisfies for any $v \in H^1_0(\Omega)$ the local error estimates*

$$h_T^{-1}\|v - R_h v\|_{L^2(T)} + \|\nabla(v - R_h v)\|_{L^2(T)} \le C\|\nabla v\|_{L^2(\omega_T)}.$$

The constant C depends on the shape-regularity and on the shapes (but not the size) of the nodal patches.

PROOF. We fix $T \in \mathcal{T}$ and $v \in H^1_0(\Omega)$. Then the error $R_h v$ restricted to T has the representation

$$R_h v|_T = \sum_{z\in\mathcal{N}(T)} (R_h v)(z)\varphi_z.$$

We use that the φ_z sum up to 1 on T and the Young inequality $(a+b+c)^2 \le 3(a^2+b^2+c^2)$ that

$$\|v - R_h v\|^2_{L^2(T)} = \|\sum_{z\in\mathcal{N}(T)} (v - (R_h v)(z))\varphi_z\|^2_{L^2(T)} \le 3\sum_{z\in\mathcal{N}(T)} \|v - (R_h v)(z)\|^2_{L^2(T)}$$

where we used $\|\varphi_z\|_{L^\infty(T)} = 1$. If $z \in \Omega$ is an interior vertex, we obtain from the Poincaré inequality that

$$\begin{aligned}\|v - (R_h v)(z)\|_{L^2(T)} \le \|v - \fint_{\omega_z} v\,dx\|_{L^2(T)} &\le \|v - \fint_{\omega_z} v\,dx\|_{L^2(\omega_z)} \\ &\le CC_P(\omega_z)h_T\|\nabla v\|_{L^2(\omega_z)}.\end{aligned}$$

If $z \in \Omega$ is a boundary vertex, then $(R_h v)(z) = 0$ and z belongs to a boundary edge, so that v has zero boundary conditions on a part of $\partial\omega_z$ of positive surface measure. Thus, Friedrichs' inequality implies

$$\|v - (R_h v)(z)\|_{L^2(T)} = \|v\|_{L^2(T)} \le \|v\|_{L^2(\omega_z)} \le h_T CC_F\|\nabla v\|_{L^2(\omega_z)}.$$

Thus, we have the local L^2 bound

$$\|v - R_h v\|_{L^2(T)} \le Ch_T\|\nabla v\|_{L^2(\omega_T)}.$$

In order to show the local bound on the gradient, we again use the above representation of R_h and the fact that the $\nabla\varphi_z$ sum up to zero on T. From this we see that

$$\begin{aligned}\|\nabla R_h v\|^2_{L^2(T)} &= \| \sum_{z\in\mathcal{N}(T)} (v - (R_h v)(z))\nabla\varphi_z\|^2_{L^2(T)} \\ &\le 3 \sum_{z\in\mathcal{N}(T)} \|v - (R_h v)(z)\|^2_{L^2(T)} \|\nabla\varphi_z\|_{L^\infty(T)} \\ &\le C h_T^{-1} \sum_{z\in\mathcal{N}(T)} \|v - (R_h v)(z)\|^2_{L^2(T)}.\end{aligned}$$

Here we used $\|\nabla\varphi_z\|_{L^\infty(T)} \le C h_T^{-1}$, see Problem 1.47. The above bounds on $v - (R_h v)(z)$ in the L^2 norms then imply the assertion. Note that the constants $C_P(\omega_z)$ and $C_F(\omega_z)$ are independent of the diameter of $C_P(\omega_z)$. They only depend on the shape of the patch. □

1.4. Linear parabolic problems

1.4.1. The heat equation and a numerical scheme. So far we have seen PDEs that were depending on spatial variables in some domain Ω. We assumed uniform positive definiteness on the diffusion coefficient A (see Theorem 1.46) so that we could use arguments based on coercivity. Such partial differential operators are called *elliptic*. We now introduce an additional time variable $t \in [0, T]$ such that the PDE is posed on the space-time cylinder $\Omega \times [0, T]$. For simplicity we will focus on the *heat equation* as a prototype. It seeks a function $u : \Omega \times [0, T] \to \mathbb{R}$ such that

$$\partial_t u - \Delta u = f \quad \text{in } \Omega \times (0, T] \to \mathbb{R}, \tag{12a}$$

$$u = 0 \quad \text{on } \partial\Omega \times [0, T], \tag{12b}$$

$$u = g \quad \text{on } \Omega \times \{0\}. \tag{12c}$$

Note that $u = u(t, x)$ and that the Laplacian Δ acts with respect to the spatial variable x. Equation (12a) is called the heat equation. It describes the time evolution of heat diffusion in the domain Ω. Condition (12b) is the homogeneous Dirichlet boundary condition on $\partial\Omega$ that holds of all times $t \in [0, T]$. Finally, (12b) is an initial condition on the initial state $u(\cdot, 0)$ that should equal a given function $g = g(x)$. The right-hand side $f = f(x, t)$ models (time-dependent) heat sources in the domain Ω.

Before stating a weak formulation for this problem, we give some brief remarks on integration of functions with values in some Banach space X. The construction is analogous to the usual Lebesgue integral.

Definition 1.63 (integral of X-valued functions). Given a Banach space $(X, \|\cdot\|)$, a function $s : [0, T] \to X$ is called a *simple function* if it has the form form $s(t) = \sum_{j=1}^m 1_{A_j}(t) u_j$ with $u_j \in X$ and Lebesgue measurable sets $A_j \subseteq [0, T]$. A function $f : [0, T] \to X$ is said to be *strongly measurable* if it is the limit (a.e. in $[0, T]$) of a sequence of simple functions. The integral of a simple function $s(t) = \sum_{j=1}^m 1_{A_j}(t) u_j$ is defined as

$$\int_0^T s(t)dt := \sum_{j=1}^m \mathrm{meas}(A_j) u_j.$$

A strongly measurable function f is said to be *summable* if there is a sequence $(s_k)_k$ of simple functions such that

$$\int_0^T \|s_k(t) - f(t)\| dt \to 0 \quad \text{as } k \to \infty.$$

For summable f we define the integral

$$\int_0^T f(t)dt := \lim_{k\to\infty} \int_0^T s_k(t)dt.$$

This method of integration is sometimes named after *S. Bochner*. ♦

We now define, for $1 \le p < \infty$,

$$L^p(0,T;X) := \{f : [0,T] \to X : f \text{ strongly measurable and } \|f\|_{L^p(0,T;X)} < \infty\}$$

where

$$\|f\|_{L^p(0,T;X)} := \left(\int_0^T \|f(t)\|^p dt \right)^{1/p}.$$

We define a weak derivative as follows.

DEFINITION 1.64. Let $v \in L^2(0,T;X)$. A summable $g : [0,T] \to X$ is called the weak derivative of v if

$$\int_0^T \partial_t \psi(t) v(t)\, dt = -\int_0^T \psi(t) g(t)\, dt \quad \text{for all } \psi \in C_c^\infty(0,T) \text{ (scalar test functions).}$$

and we write $v' = \partial_t v = g$. ♦

In many respects we can operate with Bochner integrals as with ordinary Lebesgue integrals, and this viewpoint will basically be sufficient for our lecture.
We have the following embedding of weakly differentiable functions.

LEMMA 1.65. *Let $u \in L^1(0,T;X)$ be summable with $u' \in L^1(0,T;X)$. Then $u \in C([0,T];X)$ and we have*

$$u(t) = u(s) + \int_s^t u'(r)dr \quad \textit{for all } 0 \le s \le t \le T.$$

PROOF. We only sketch the idea of the proof. As in prior sections on Sobolev spaces, we can approximate u by some u_ε (through convolution) and see that $u_\varepsilon \to u$ in $L^1(0,T;X)$ as well as $u_\varepsilon \to u'$ on $L^1(s,t;X)$ for compact intervals $[s,t] \subseteq (0,T)$ as $\varepsilon \to 0$. We observe

$$u_\varepsilon(t) = u_\varepsilon(s) + \int_s^t u_\varepsilon'(r)dr$$

and pass to the limit $\varepsilon \to 0$, which shows

$$u(t) = u(s) + \int_s^t u'(r)dr$$

for almost every $0 < s < t < T$. From this representation we see that u is continuous because the integral is continuous as a function of t. □

REMARK 1.66. The statement that $u \in L^1(0,T;X)$ is continuous should always be read as: There exists a continuous function in the equivalence class u. ♦

Recall the dual space $H^{-1}(\Omega)$ of $H_0^1(\Omega)$.

LEMMA 1.67. *If $u \in L^2(0,T;H_0^1(\Omega))$ with $u' \in L^2(0,T;H^{-1}(\Omega))$ be given. Then $u \in C([0,T];L^2(\Omega))$ and furthermore*

$$\partial_t \|u(t)\|_{L^2(\Omega)}^2 = 2\langle u'(t), u(t)\rangle.$$

PROOF. The proof again works by regularizing u and showing convergence of the regularization u_ε in $C([0,T];L^2(\Omega)$. The claimed formula for the derivative of the squared norm is easily verified for smooth functions u_ε and remains true after taking limits. The details can be found in [**Eva10**, §5.9]. □

Let us derive a weak formulation for (12). We interpret Δ as the weak Laplacian on $H^1_0(\Omega)$. If we assume $f \in L^2(\Omega \times [0,T])$ for the right-hand side, the solution $u(\cdot,t)$ belongs to $H^1_0(\Omega)$ for all $t \in [0,T]$. We thus have $u \in L^2([0,T];H^1_0(\Omega))$. From (12a), we see that $\partial_t u$ equals $\Delta u + f$ at all times, which is an element of $H^{-1}(\Omega)$.

DEFINITION 1.68. Let $\Omega \subseteq \mathbb{R}^2$ be an open, bounded, connected Lipschitz polygon and $0 < T < \infty$. Let $f \in L^2(\Omega \times [0,T])$ and $g \in L^2(\Omega)$ be given. A function $u \in L^2([0,T];H^1_0(\Omega))$ with $\partial_t u \in L^2([0,T];H^{-1}(\Omega))$ is said to be a *solution to the initial/boundary-value problem* of the heat equation if it satisfies

$$\langle \partial_t u(\cdot,t), v\rangle + \int_\Omega \nabla u(\cdot,t)\cdot\nabla v\,dx = \int_\Omega f(\cdot,t)v\,dx \quad \text{for all } v \in H^1_0(\Omega) \text{ and a.e. } t \in [0,T]$$

and

$$u(\cdot,0) = g.$$

From Lemma 1.65 we see that posing a condition on $u(\cdot,0)$ is meaningful. ♦

Numerical methods. We will later prove that there exists a unique solution to the heat equation. Let us first define a numerical method, so that we can start to do actual computations. The idea is to discretize the spatial derivatives with a finite element method. The time derivative is discretized by difference quotients. Ignoring for the moment the time discretization, we use finite elements (with respect to a triangulation $\mathcal{T}$ of Ω) in space and obtain the so-called *semidiscrete* equation: Seek $\tilde{u}_h \in S^1_0(\mathcal{T}) \times [0,T]$ such that

$$\langle \partial_t \tilde{u}_h(\cdot,t), v_h\rangle + \int_\Omega \nabla \tilde{u}_h(\cdot,t)\cdot\nabla v_h\,dx = \int_\Omega f(\cdot,t)v_h\,dx \quad \text{for all } v_h \in S^1_0(\mathcal{T}_h) \text{ and a.e. } t \in [0,T] \tag{13}$$

and

$$\tilde{u}_h(\cdot,0) = g_h$$

where g_h is a suitable approximation to g, for instance the L^2 projection to $S^1_0(\mathcal{T})$ or some (quasi-)interpolation.
The equation is called semidiscrete because the dependence on time has not been resolved by a numerical method, yet. In order to obtain an actual numerical method, we need to discretize the semidiscrete problem (13) in time. To this end, we approximate the time derivative by difference quotients.

DEFINITION 1.69. Given a time step size Δt and a sequence $(U_j)_{j=0,\dots,J}$ of elements of some vector space, we define

$$\partial_t^+ U_j := \frac{U_{j+1}-U_j}{\Delta t}, \quad (j = 0,\dots,J-1) \quad \textit{(forward difference quotient)}$$

and

$$\partial_t^- U_j := \frac{U_j - U_{j-1}}{\Delta t}, \quad (j = 1,\dots,J) \quad \textit{(backward difference quotient)}.$$

♦

Let now the interval $[0,T]$ be uniformly subdivided by the time step size $\Delta t = T/J$ as

$$t_0 = 0,\ t_1 = \Delta t,\ \dots,\ t_j = j\Delta t,\ \dots,\ t_J = T.$$

By Taylor expansion one derives the following approximation property.

LEMMA 1.70. *Given $u \in C^2([0,T])$, we have for ∂_t^+ and ∂_t^- that*

$$|\partial_t^{\pm} u(t_j) - \partial_t u(t_j)| \leq \frac{\Delta t}{2} \|\partial_{tt}^2 u\|_{C([0,T])}.$$

PROOF. Problem 1.49. □

We introduce a uniform time step size $\Delta t = 1/J$. If we replace the time derivative in (13) by the backward difference quotient ∂_t^-, we arrive at fully discrete problem. We denote by $(u_h^k)_{k=0}^J$ the sequence of spatial unknowns in $S_0^1(\mathcal{T})$ and obtain the equations

$$\langle \partial_t^- u_h^k, v_h \rangle + \int_\Omega \nabla u_h^k \cdot \nabla v_h \, dx = \int_\Omega f(\cdot, t_k) v_h \, dx \quad \text{for all } v_h \in S_0^1(\mathcal{T}) \text{ and all } k = 1, \ldots, J. \tag{14}$$

The initial condition is $u_h^0 = g_h$ where g_h is some approximation to g in $S_0^1(\mathcal{T})$. In matrix-vector notation for the coefficient vector x^k of u_h^k this reads as

$$M^\top \partial_t^- x^k + A^\top x^k = b_k$$

where M is the mass matrix, A is the stiffness matrix, and b_k is the right-hand side vector for f at t_k. We use the definition of ∂_t^- and the fact that the x^0 is known and obtain the following numerical scheme

$$(M^\top + \Delta t A^\top) x_k = \Delta t b_k + M^\top x^{k-1} \quad \text{for } j = 1, \ldots, J.$$

This method is called the *implicit Euler method* or *backward Euler method*. We note that in each time step we have to solve a linear system.
We can perform an analogous derivation for the forward difference quotient ∂_t^+ and obtain the following scheme

$$M^\top x_k = \Delta t b_k + (M^\top - \Delta t A^\top) x_{k-1} \quad \text{for } j = 1, \ldots, J$$

called the *explicit Euler method* or *forward Euler method*. Here, a system with the mass matrix M has to be solved in each step. It turns out that this can be done much more efficiently in comparison with the implicit Euler method. The reason is that the mass matrix can be suitably approximated by the diagonal matrix $\tilde{M}$ whose entries M_{jj} equals the row sum of M. This procedure is referred to as *mass lumping*. Each time step in the lumped scheme is then very cheap because no linear system has to be solved. This explains why the scheme is called *explicit*. It will, however, turn out that the time step size needs to be chosen much smaller for the explicit method than for the implicit method in order to obtain reasonable approximations.

1.4.2. Error analysis. We now perform an error analysis for the implicit Euler method. The proof follows a general guideline for the error analysis of any time-stepping scheme. The two building blocks are:

- *Stability:* The sequence of discrete approximations stays bounded (uniformly in Δt) in norms that we expect to be bounded for the exact solution. This is a reasonable requirement for convergence with respect to these norms.
- *Consistency:* Usually, the exact (or semidiscrete) solution will not satisfy the recursion rule of the numerical method. Consistency means that the resulting error terms converge to zero for $\Delta t \to 0$.

LEMMA 1.71 (stability). *The iterates u_h^k of the backward Euler scheme satisfy*

$$\max_{k=1,\ldots,J} \|u_h^k\|_{L^2(\Omega)}^2 + \Delta t \sum_{k=1}^J \|\nabla u_h^k\|_{L^2(\Omega)}^2 \leq 2\|u_h^0\|_{L^2(\Omega)}^2 + 2\Delta t \sum_{k=1}^J \|f(\cdot, t_k)\|_{H^{-1}(\Omega)}^2.$$

PROOF. From the definition of ∂_t^- we obtain the following identity

$$u_h^k = \frac{1}{2}(u_h^k + u_h^{k-1}) + \frac{1}{2}\Delta t \partial_t^- u_h^k.$$

A straightforward computation then yields

$$\begin{aligned}\langle \partial_t^- u_h^k, u_h^k \rangle = \int_\Omega \partial_t^- u_h^k u_h^k \, dx &= \int_\Omega \partial_t^- u_h^k (\frac{1}{2}(u_h^k + u_h^{k-1})) \, dx + \int_\Omega \partial_t^- u_h^k (\frac{1}{2}\Delta t \partial_t^- u_h^k) \, dx \\ &= \frac{1}{2\Delta t}(\|u_h^k\|_{L^2(\Omega)}^2 - \|u_h^{k-1}\|_{L^2(\Omega)}^2) + \frac{1}{2}\Delta t \|\partial_t^- u_h^k\|_{L^2(\Omega)}^2.\end{aligned}$$

We use $v_h := u_h^k$ as a test function at step k of (14) and obtain

$$\begin{aligned}&\frac{1}{2\Delta t}(\|u_h^k\|_{L^2(\Omega)}^2 - \|u_h^{k-1}\|_{L^2(\Omega)}^2) + \frac{1}{2}\Delta t \|\partial_t^- u_h^k\|_{L^2(\Omega)}^2 + \|\nabla u_h^k\|_{L^2(\Omega)}^2 \\ &\qquad = \int_\Omega f(\cdot, t_k) u_h^k \, dx \le \|f(\cdot, t_k)\|_{H^{-1}(\Omega)} \|\nabla u_h^k\|_{L^2(\Omega)} \\ &\qquad \le \frac{1}{2}\|f(\cdot, t_k)\|_{H^{-1}(\Omega)}^2 + \frac{1}{2}\|\nabla u_h^k\|_{L^2(\Omega)}^2.\end{aligned}$$

After re-arranging the gradient terms and estimating the norm of ∂_t^- from below by 0, multiplication of the estimate by $2\Delta t$ and summation over k results in

$$\sum_{k=1}^K (\|u_h^k\|_{L^2(\Omega)}^2 - \|u_h^{k-1}\|_{L^2(\Omega)}^2) + \Delta t \sum_{k=1}^K \|\nabla u_h^k\|_{L^2(\Omega)}^2 \le \Delta t \sum_{k=1}^K \|f(\cdot, t_k)\|_{H^{-1}(\Omega)}^2$$

for any $K \le J$. The first term is a telescoping sum and equals $(\|u_h^K\|_{L^2(\Omega)}^2 - \|u_h^0\|_{L^2(\Omega)}^2)$. Increasing the right-hand side by replacing K by J, we thus see that

$$\max_{k=1,\dots,J} \|u_h^k\|_{L^2(\Omega)}^2 \le \|u_h^0\|_{L^2(\Omega)}^2 + \Delta t \sum_{k=1}^J \|f(\cdot, t_k)\|_{H^{-1}(\Omega)}^2.$$

The combination with the foregoing estimate then implies the assertion. □

The solution of the spatial finite element method defines a map $G_h : H_0^1(\Omega) \to S_0^1(\mathcal{T})$, called the *Galerkin projection*. The Galerkin orthogonality reads

$$\int_\Omega \nabla(u - G_h u) \cdot \nabla v_h \, dx = 0 \quad \text{for all } v_h \in S_0^1(\mathcal{T}).$$

On convex domains, we have from elliptic regularity (Theorem 1.60 and Corollary 1.59) that

$$\|\nabla(u - G_h u)\|_{L^2(\Omega)} \le Ch \|D^2 u\|_{L^2(\Omega)}. \tag{15a}$$

Furthermore, Theorem 1.61 implies

$$\|u - G_h u\|_{L^2(\Omega)} \le Ch^2 \|D^2 u\|_{L^2(\Omega)}. \tag{15b}$$

LEMMA 1.72 (consistency). *Assume that the domain Ω is convex and that the solution u to the heat equation additionally satisfies*

$$u \in C^1([0,T]; H^2(\Omega)) \cap C^2([0,T]; L^2(\Omega)).$$

Then, the Galerkin projection $z_h^k := G_h u(\cdot, t_k)$ of the exact solution $u(\cdot, t_k)$ at t_k satisfies, for all $k = 1, \dots, J$,

$$\langle \partial_t^- z_h^k, v_h \rangle + \int_\Omega \nabla z_h^k \cdot \nabla v_h \, dx = \int_\Omega f(\cdot, t_k) v_h \, dx + \mathcal{C}_k(v_h)$$

where the $\mathcal{C}_k \in H^{-1}(\Omega)$ are linear functionals satisfying

$$\Delta t \sum_{k=1}^J \|\mathcal{C}_k\|_{H^{-1}(\Omega)}^2 \le C(h^4 \int_0^T \|D^2 \partial_t u(\cdot, s)\|_{L^2(\Omega)}^2 \, ds + (\Delta t)^2 \int_0^T \|\partial_{tt} u(\cdot, s)\|_{L^2(\Omega)}^2 \, ds).$$

PROOF. We have the Galerkin orthogonality

$$\int_\Omega \nabla(z_h^k - u(\cdot,t_k)) \cdot \nabla v_h \, dx = 0.$$

This and the solution property of $u(\cdot,t_k)$ show

$$\begin{aligned}
&\langle \partial_t^- z_h^k, v_h\rangle + \int_\Omega \nabla z_h^k \cdot \nabla v_h \, dx \\
&= \langle \partial_t^- z_h^k - \partial_t u(\cdot,t_k), v_h\rangle + \langle \partial_t u(\cdot,t_k), v_h\rangle + \int_\Omega \nabla u(\cdot,t_k)\cdot\nabla v_h \, dx \\
&= \mathcal{C}_k + \int_\Omega f(\cdot,t_k) v_h \, dx
\end{aligned}$$

where $\mathcal{C}_k$ is defined by

$$\mathcal{C}_k(v) := \int_\Omega (\partial_t^- z_h^k - \partial_t u(\cdot,t_k)) v dx \quad \text{for any } v \in H_0^1(\Omega).$$

To estimate the H^{-1} norm of $\mathcal{C}_k$, we split the consistency term on the right-hand side as follows

$$\mathcal{C}_k(v) = \int_\Omega \partial_t^-(z_h^k - u(\cdot,t_k)) v dx + \int_\Omega (\partial_t^- u(\cdot,t_k) - \partial_t u(\cdot,t_k)) v dx. \tag{16}$$

By the fundamental theorem of calculus (see also Lemma 1.65) the first term on the right-hand side of (16) equals

$$\begin{aligned}
\int_\Omega \partial_t^-(z_h^k - u(\cdot,t_k)) v dx &= \frac{1}{\Delta t}\int_{t_{k-1}}^{t_k}\int_\Omega (G_h \partial_t u(\cdot,s) - \partial_t u(\cdot,s)) v \, dx \, ds \\
&\le \frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} \|G_h\partial_t u(\cdot,s) - \partial_t u(\cdot,s)\|_{L^2(\Omega)} \, ds \|v\|_{L^2(\Omega)}.
\end{aligned}$$

Thus we obtain with the error bound (15a) on the Galerkin projection and Hölder's inequality that

$$\begin{aligned}
\int_\Omega \partial_t^-(z_h^k - u(\cdot,t_k)) v dx &\le C\frac{h^2}{\Delta t}\int_{t_{k-1}}^{t_k} \|D^2\partial_t u(\cdot,s)\|_{L^2(\Omega)} \, ds \|v\|_{L^2(\Omega)} \\
&\le C\frac{h^2}{\sqrt{\Delta t}}\sqrt{\int_{t_{k-1}}^{t_k} \|D^2\partial_t u(\cdot,s)\|_{L^2(\Omega)}^2 \, ds}\|v\|_{L^2(\Omega)}.
\end{aligned}$$

For the difference in the second term on the right-hand side of (16), we obtain through Taylor's formula

$$\begin{aligned}
\partial_t^- u(\cdot,t_k) - \partial_t u(\cdot,t_k) &= \frac{1}{\Delta t}(u(\cdot,t_k) - u(\cdot,t_{k-1})) - \partial_t u(\cdot,t_k) \\
&= -\frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} (s - t_{k-1}) \partial_{tt} u(\cdot,s) \, ds.
\end{aligned}$$

Thus, with Hölder's inequality,

$$\begin{aligned}
\int_\Omega (\partial_t^- u(\cdot,t_k) - \partial_t u(\cdot,t_k)) v_h dx &= -\frac{1}{\Delta t}\int_{t_{k-1}}^{t_k} (s - t_{k-1}) \int_\Omega \partial_{tt} u(\cdot,s) v_h \, dx \, ds \\
&\le \int_{t_{k-1}}^{t_k} \|\partial_{tt} u(\cdot,s)\|_{L^2(\Omega)} \, ds \|v_h\|_{L^2(\Omega)} \\
&\le \sqrt{\Delta t}\sqrt{\int_{t_{k-1}}^{t_k} \|\partial_{tt} u(\cdot,s)\|_{L^2(\Omega)}^2 \, ds}\|v_h\|_{L^2(\Omega)}.
\end{aligned}$$

Combining the foregoing estimates with (16) results in

$$\mathcal{C}_k(v_h) \le \left(C\frac{h^2}{\sqrt{\Delta t}} \sqrt{\int_{t_{k-1}}^{t_k} \|D^2 \partial_t u(\cdot,s)\|_{L^2(\Omega)}^2 \, ds} \right. \\ \left. + \sqrt{\Delta t} \sqrt{\int_{t_{k-1}}^{t_k} \|\partial_{tt} u(\cdot,s)\|_{L^2(\Omega)}^2 \, ds} \right) \|v_h\|_{L^2(\Omega)}.$$

This implies a bound on $\|\mathcal{C}_k\|_{H^{-1}}$. Taking squares and summing over k implies the stated bound on the sum of $\|\mathcal{C}_k\|_{H^{-1}}^2$. □

THEOREM 1.73 (error estimate for the implicit Euler method). *Under the assumption of Lemma 1.72, the implicit Euler method with initial value $u_h^0 = I_h g$ (nodal interpolation) has the approximation order*

$$\sqrt{\Delta t \sum_{k=1}^{J} \|\nabla(u(\cdot,t_k) - u_h^k)\|_{L^2(\Omega)}^2} \le \mathcal{O}(h + \Delta t)$$

and

$$\max_{k=1,\dots,J} \|u(\cdot,t_k) - u_h^k\|_{L^2(\Omega)} \le \mathcal{O}(h^2 + \Delta t).$$

PROOF. We introduce the Galerkin projections $z_k := G_h u(\cdot,t_k)$ and split the error as follows

$$\Delta t \sum_{k=1}^{J} \|\nabla(u(\cdot,t_k) - u_h^k)\|_{L^2(\Omega)}^2 \\ \le 2\Delta t \sum_{k=1}^{J} \|\nabla(u(\cdot,t_k) - z_k)\|_{L^2(\Omega)}^2 + 2\Delta t \sum_{k=1}^{J} \|\nabla(z_k - u_h^k)\|_{L^2(\Omega)}^2.$$

The first term is estimated by the approximation property (15a) of the Galerkin projection and $\Delta T = 1/J$ as follows

$$2\Delta t \sum_{k=1}^{J} \|\nabla(u(\cdot,t_k) - z_k)\|_{L^2(\Omega)}^2 \le Ch^2 \Delta t \sum_{k=1}^{J} \|D^2 u(\cdot,t_k)\|_{L^2(\Omega)}^2 \\ \le Ch^2 \|u\|_{C^0([0,T];H^2(\Omega))}.$$

For the second term, observe that Lemma 1.72 implies that $z_k - u_h^k$ is the sequence of an implicit Euler scheme with right-hand side $\mathcal{C}_k$. Therefore, the stability of Lemma 1.71 shows

$$\Delta t \sum_{k=1}^{J} \|\nabla u_h^k - z_k\|_{L^2(\Omega)}^2 \le 2\|I_h g - z_0\|_{L^2(\Omega)} + 2\Delta t \sum_{k=1}^{J} \|\mathcal{C}_k\|_{H^{-1}(\Omega)} \\ \le 2\|g - G_h g\|_{L^2(\Omega)} + Ch^4 \int_0^T \|D^2 \partial_t u(\cdot,s)\|_{L^2(\Omega)}^2 \, ds \\ + C(\Delta t)^2 \int_0^T \|\partial_{tt} u(\cdot,s)\|_{L^2(\Omega)}^2 \, ds.$$

Since we have

$$\|I_h g - z_0\|_{L^2(\Omega)} \le \|I_h g - g\|_{L^2(\Omega)} + \|g - z_0\|_{L^2(\Omega)} \le Ch^2 \|D^2 g\|_{L^2(\Omega)}$$

(note that $g \in H^2(\Omega)$), the error estimate for the norm involving the gradient is established. The error estimate for the maximal L^2 error is shown analogously and left as an exercise. □

We have seen in Theorem 1.73 that the choice $h \approx \Delta t$ yields a balanced error bound for the discrete L^2-H^1 norm. This is the case for the implicit Euler method. The explicit Euler method satisfies a similar error bound under more restrictive assumptions. A stability analysis of the explicit Euler method shows that stability is achieved under the additional condition

$$\frac{\Delta t}{h^2} \leq c$$

for some global constant c. This means that we have to chose the time step much smaller, namely of order $(\Delta t)^2$, which is not rewarded by the error estimate. In spite of the low computational costs in each time step, this makes the explicit method rather unattractive. In contrast, the implicit scheme is *unconditionally stable*.

1.4.3. Existence and uniqueness for the heat equation. We now prove existence of weak solutions in a constructive procedure. We follow the following roadmap. In a first step, we discretize the PDE in space with a finite-dimensional Galerkin method. At this stage, we are not interested in actual numerical computations but rather use the Galerkin method as a tool from analysis. The space-discretized PDE can then be interpreted as a system of ordinary differential equations, which is solvable by known arguments. In a second step, we derive so-called *energy estimates* stating that certain norms of the space-discrete solutions are uniformly bounded with respect to the dimension of our Galerkin subspace. In the third step, we pass to the limit and see that the Galerkin solutions weakly converge to some limit, which is then proven to satisfy the heat equation.
In order to define Galerkin approximations, let $\mathcal{T}_0$ be a triangulation of Ω. We consider the shape-regular sequence $(\mathcal{T}_j)_j$ of triangulations resulting from j red refinements. The straight-forward space-discrete Galerkin method (already introduced in (13)) is to find $u_j \in S^1_0(\mathcal{T}_j) \times [0,T]$ such that

$$\langle \partial_t u_j(\cdot,t), v_j \rangle + \int_\Omega \nabla u_j(\cdot,t) \cdot \nabla v_j \, dx = \int_\Omega f(\cdot,t) v_j \, dx \qquad \text{for all } v_j \in S^1_0(\mathcal{T}_j) \text{ and all } t \in [0,T]$$

and

$$u_j(\cdot,0) = \Pi^{(j)} g.$$

Where $\Pi^{(j)} g$ is the L^2 projection of g to the finite element space $S^1_0(\mathcal{T}_j)$.

LEMMA 1.74. *For each $j = 0, 1, \ldots$ there exists a unique (semidiscrete) Galerkin approximation u_j to the heat equation.*

PROOF. Letting $x_j(t)$ denote the coefficient vector of the spatial part of u_j, we see that the Galerkin equation is equivalent to

$$M^\top \partial_t x_j(t) + A^\top x_j(t) = b(t)$$

where M is the mass matrix, A is the stiffness matrix, and $b(t)$ is the right-hand side vector. This is a linear ODE system which is complemented by the initial condition $x_j(0) = y_j$ where y_j are the coefficients of $\Pi^{(j)} g$. Thus, it is uniquely solvable by standard results on ODEs. □

We now turn to the announced energy estimates.

THEOREM 1.75 (energy estimates). *There exists a constant $C > 0$ (independent of f, g, j, u_j) such that, for all $j = 0, 1, \ldots$,*

$$\max_{0 \leq t \leq T} \|u_j(\cdot,t)\|_{L^2(\Omega)} + \|\nabla u_j\|_{L^2([0,T];L^2(\Omega))} + \|\partial_t u\|_{L^2([0,T];H^{-1}(\Omega))} \leq C(\|f\|_{L^2([0,T];L^2(\Omega))} + \|g\|_{L^2(\Omega)}.)$$

PROOF. An application of the chain rule shows the identity

$$\langle \partial_t u_j, u_j \rangle = \partial_t (\frac{1}{2} \|u_j\|^2_{L^2(\Omega)}).$$

We use the test function $v_j = u_j(t)$ in the Galerkin equation, multiply by 2 and obtain with the Cauchy and Young inequalities

$$\partial_t (\|u_j\|^2_{L^2(\Omega)}) + 2\|\nabla u_j\|^2_{L^2(\Omega)} = 2 \int_\Omega f u_j \, dx \leq \|f\|^2_{L^2(\Omega)} + \|u_j\|^2_{L^2(\Omega)} \tag{17}$$

for almost every $t \in [0,T]$. This is a differential inequality bounding the growth of $\|u_j(\cdot,t)\|^2_{L^2(\Omega)}$ by the quantity itself and $\|f(\cdot,t)\|^2_{L^2(\Omega)}$. Gronwall's lemma thus implies the bound

$$\|u_j(\cdot,t)\|^2_{L^2(\Omega)} \leq \exp(t) \left(\|u_j(\cdot,0)\|^2_{L^2(\Omega)} + \int_0^t \|f(\cdot,s)\|^2_{L^2(\Omega)} ds \right).$$

Note that $u_j(\cdot,0)$ equals the L^2 projection of g, whence $\|u_j(\cdot,0)\|_{L^2(\Omega)} \leq \|g\|_{L^2(\Omega)}$. We then have (with $C_1 = \exp(T)$)

$$\begin{aligned} \max_{0\leq t\leq T} \|u_j(\cdot,t)\|^2_{L^2(\Omega)} &\leq C_1 (\|f\|^2_{L^2([0,T];L^2(\Omega))} + \|g\|^2_{L^2(\Omega)}) \\ &\leq C_1 (\|f\|_{L^2([0,T];L^2(\Omega))} + \|g\|_{L^2(\Omega)})^2 \end{aligned}$$

and thus the bound for the first term on the left-hand side of the assertion.
We now integrate (17) with respect to time and deduce

$$\begin{aligned} &\frac{1}{2} \|\nabla u_j\|^2_{L^2([0,T];\Omega)} \\ &\quad \leq \|f\|^2_{L^2([0,T];\Omega)} + \int_0^T \|u_j(\cdot,s)\|^2_{L^2(\Omega)} ds + \|u_j(\cdot,0)\|^2_{L^2(\Omega)} - \|u_j(\cdot,T)\|^2_{L^2(\Omega)}. \end{aligned}$$

With the bound on the maximum on $\|u_j(\cdot,s)\|_{L^2(\Omega)}$ just shown we thus find

$$\begin{aligned} \frac{1}{2} \|\nabla u_j\|^2_{L^2([0,T];\Omega)} &\leq \|f\|^2_{L^2([0,T];\Omega)} + (T+2) \max_{0\leq t\leq T} \|u_j(\cdot,t)\|^2_{L^2(\Omega)} \\ &\leq (1 + C_1(T+2)) (\|f\|_{L^2([0,T];\Omega)} + \|g\|_{L^2(\Omega)})^2 \end{aligned}$$

which implies the bound on the second term on the left-hand side of the asserted estimate.
In order to bound the third term on the left-hand side of the assertion including the negative norm, let $v \in H^1_0(\Omega)$ with $\|\nabla v\|_{L^2(\Omega)} = 1$ be arbitrary. We denote by $\Pi^{(j)} v \in S^1_0(\mathcal{T}_j)$ the L^2 projection of v to the finite element space and use it to test the Galerkin equation. We obtain

$$\langle \partial_t u_j(\cdot,t), \Pi^{(j)} v \rangle + \int_\Omega \nabla u_j(\cdot,t) \cdot \nabla \Pi^{(j)} v \, dx = \int_\Omega f(\cdot,t) \Pi^{(j)} v \, dx.$$

Note that the term with angular brackets on the left-hand side is nothing but the L^2 product because the solution to the ODE system is the FEM function with coefficients $\partial_t x_j$. In the L^2 inner products (without gradients) in the above identity, the projection property of the L^2 projection shows that we can replace the function $\Pi^{(j)} v$ by v. After rearranging terms we therefore have

$$\langle \partial_t u_j(\cdot,t), v \rangle = \int_\Omega f(\cdot,t) \Pi^{(j)} v \, dx - \int_\Omega \nabla u_j(\cdot,t) \cdot \nabla \Pi^{(j)} v \, dx.$$

With the Cauchy-Schwarz inequality and the nonexpansivity of $\Pi^{(j)}$ with respect to the L^2 norm we then obtain the bound

$$|\langle \partial_t u_j, v \rangle| \leq \|f\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)} + \|\nabla u_j\|_{L^2(\Omega)} \|\nabla \Pi^{(j)} v\|_{L^2(\Omega)}.$$

We have seen in Problem 1.52 that the L^2 projection is H^1 stable on sequences of red refined meshes so that there is some constant C_3 with $\|\nabla \Pi^{(j)} v\|_{L^2(\Omega)} \leq C_3 \|\nabla v\|_{L^2(\Omega}$. We use this bound and the Friedrichs inequality for $\|v\|_{L^2(\Omega)}$ in the above estimate and get

$$|\langle \partial_t u_j, v\rangle| \leq C_4(\|f\|_{L^2(\Omega)} + \|\nabla u_j\|_{L^2(\Omega)})$$

for some constant C_4 because $\|\nabla v\|_{L^2(\Omega)}$. Taking the supremum over such v and integrating we obtain

$$\int_0^T \|\partial_t u_j\|_{H^{-1}(\Omega)} \leq C_4(\|f\|^2_{L^2([0,T];\Omega)} + \|\nabla u_j\|^2_{L^2([0,T];\Omega)}\cdot).$$

We can now use the (already established) bound on $\|\nabla u_j\|_{L^2([0,T];\Omega)}$ to prove the estimate on the third term on the left-hand side of the asserted inequality. □

THEOREM 1.76 (existence of a weak solution). *There exists a weak solution to the initial/boundary value problem of the heat equation.*

PROOF. The bounds from the energy estimates show that the sequence $(u_j)_j$ of Galerkin solutions is bounded in $L^2([0,T]; H^1_0(\Omega))$ and $(\partial_t u_j)_j$ is bounded in $L^2([0,T]; H^{-1}(\Omega))$. Since these spaces are reflexive, there is a subsequence (which we do not relabel with an additional index) and some $u \in L^2([0,T]; H^1_0(\Omega))$ and $v \in L^2([0,T]; H^{-1}(\Omega))$ such that

$$\begin{cases} u_j \rightharpoonup u & \text{weakly in } L^2([0,T]; H^1_0(\Omega)) \\ \partial_t u_j \rightharpoonup v & \text{weakly in } L^2([0,T]; H^{-1}(\Omega)). \end{cases}$$

It is then an exercise (cf. the arguments in the proof of completeness of Sobolev spaces) to prove that $v = \partial_t u$. The plan of the proof is to show that u satisfies the heat equation and the initial condition. We momentarily fix an index m and choose a finite element test function $v_m \in L^2([0,T]; S^1_0(\mathcal{T}_m))$. Recall that $S^1_0(\mathcal{T}_m) \subseteq S^1_0(\mathcal{T}_j)$ for any $j \geq m$ and so v_m is an admissible test function on all finer triangulations. We therefore obtain from the Galerkin equation and integration with respect to time that

$$\int_0^T \langle \partial_t u_j, v_m\rangle dt + \int_0^T \int_\Omega \nabla u_j \cdot \nabla v_m \, dxdt = \int_0^T \int_\Omega f v_m \, dxdt \quad \text{for all } j \geq m. \tag{18}$$

By the above weak convergence results, we can pass to the limit $j \to \infty$ to get

$$\int_0^T \langle \partial_t u, v_m\rangle dt + \int_0^T \int_\Omega \nabla u \cdot \nabla v_m \, dxdt = \int_0^T \int_\Omega f v_m \, dxdt. \tag{19}$$

This identity is valid for all $m \in \mathbb{N}$. Since the finite element functions on a sequence of red refined meshes are dense in $H^1_0(\Omega)$ (see Problem 1.53), the identity even holds for all $v \in L^2([0,T]; H^1_0(\Omega))$. In particular, the weak heat equation is satisfied for almost every t and all test functions $v \in H^1_0(\Omega)$.

We now proceed by showing that u satisfies the initial condition $u(\cdot, 0) = g$. We consider (18) with $v_m \in C^1([0,T]; S^1_0(\mathcal{T}_m))$ with $v_m(T) = 0$ being a test function for the Galerkin equation with $j \geq m$. Integration by parts (with respect to time) reveals

$$\int_0^T -\langle \partial_t v_m, u_j\rangle dt + \int_0^T \int_\Omega \nabla u_j \cdot \nabla v_m \, dxdt = \int_0^T \int_\Omega f v_m \, dxdt + \int_\Omega u_j(\cdot,0) v_m(\cdot,0) \, dx.$$

Letting $j \to \infty$ and observing that $u_j(\cdot,0) = \Pi^{(j)} g \to g$ in $L^2(\Omega)$, we use the weak convergence relations and find

$$\int_0^T -\langle \partial_t v, u\rangle dt + \int_0^T \int_\Omega \nabla u \cdot \nabla v \, dxdt = \int_0^T \int_\Omega f v \, dxdt + \int_\Omega g v(\cdot,0) \, dx$$

where we again used density of the finite element functions v_m. On the other hand, a similar argument for (19) results in

$$\int_0^T -\langle \partial_t v, u\rangle dt + \int_0^T \int_\Omega \nabla u \cdot \nabla v \, dx dt = \int_0^T \int_\Omega f v \, dx dt + \int_\Omega u(\cdot,0) v(\cdot,0)\, dx.$$

Comparing these two formulas then leads to $u(0) = g$ because the test function v was arbitrary. □

THEOREM 1.77 (uniqueness). *The weak solution to the initial/boundary value problem of the heat equation is unique.*

PROOF. The difference e of two weak solutions satisfies the heat equation with right-hand side $f = 0$ and initial values $g = 0$. We then have (cf. Lemma 1.67) that

$$\partial_t(\frac{1}{2}\|e\|^2_{L^2(\Omega)}) + \|\nabla e\|^2_{L^2(\Omega)} = \langle \partial_t e, e\rangle + \int_\Omega \nabla e \cdot \nabla e \, dx = 0$$

for almost all t. In particular, we have $\partial_t(\|e\|^2_{L^2(\Omega)}) \le 0$ and the norm of e is nonincreasing. The initial condition thus shows that $e = 0$. □

1.A. Problems

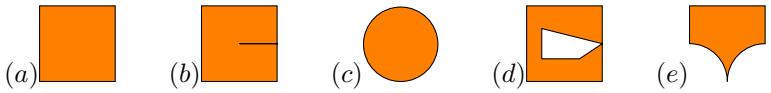
PROBLEM 1.1. Let the following function be given

$$\Phi(x) = \begin{cases} -\frac{1}{2\pi}\log|x| & \text{if } n = 2 \\ \frac{1}{n(n-2)\alpha(n)}\frac{1}{|x|^{n-2}} & \text{if } n \ge 2. \end{cases}$$

Here, $\alpha(n) \neq 0$ is some real number. Show that $\Delta\Phi(x) = 0$ holds for all $x \in \mathbb{R}^n \backslash \{0\}$.

PROBLEM 1.2. Prove, based on the divergence theorem, the formula of integration by parts as well as Green's formula.

PROBLEM 1.3. Which of the following domains possess a Lipschitz boundary?



The lower part of the boundary of (e) is parametrized by $y = \sqrt{|x|}$.

PROBLEM 1.4. Prove that the Laplacian is represented in polar coordinates (r, φ) as follows

$$\Delta = \frac{\partial^2}{\partial r^2} + \frac{1}{r}\frac{\partial}{\partial r} + \frac{1}{r^2}\frac{\partial^2}{\partial \varphi^2}.$$

PROBLEM 1.5. Verify the statements from Example 1.13.

PROBLEM 1.6. Prove the fundamental lemma of calculus of variations.

PROBLEM 1.7. Show that the weak derivative is unique. (Hint: fundamental lemma of calculus of variations)

PROBLEM 1.8. Show that the function $v(x) = \log(|\log(|x|)|)$ on the unit disk $\Omega = \{x \in \mathbb{R}^2 : |x| < 1\}$ is weakly differentiable but neither bounded nor continuous on Ω.

PROBLEM 1.9. Show that the notions of classical and weak derivative coincide for continuously differentiable functions.

PROBLEM 1.10. Draw a regular triangulation of the square $(0,1)^2$ with 7 triangles.

PROBLEM 1.11. Let K, T be triangles that intersect in one point $z = T \cap K$. The point z is vertex to T but not to K. Such point is called a *hanging node*. Draw a picture of this situation and convince yourself that regular triangulations cannot contain any hanging node.

PROBLEM 1.12. Prove the assertions from Example 1.19. Draw plots of such piecewise affine function for some examples.

PROBLEM 1.13. Prove the claims from Example 1.18.

PROBLEM 1.14. Is the sign function from (2) weakly differentiable?

PROBLEM 1.15. Show that the nodal basis $(\varphi_z)_{z\in\mathcal{N}}$ forms a partition of unity.

PROBLEM 1.16. Show that the functions φ_z are uniquely defined by (3) and that they form as basis of $S^1(\mathcal{T})$. Draw the graph of one of the basis functions φ_z on an example triangulation.

PROBLEM 1.17 (barycentric coordinates). Let $T \subseteq \mathbb{R}^2$ be a triangle with vertices z_1, z_2, z_3. Show that to any point $x \in T$ there exist unique real numbers $\lambda_1(x)$, $\lambda_2(x)$, $\lambda_3(x)$ with the properties

$$x = \lambda_1(x)z_1 + \lambda_2(x)z_2 + \lambda_3(x)z_3 \qquad \text{and} \qquad \lambda_1(x) + \lambda_2(x) + \lambda_3(x) = 1.$$

The λ_j are called *barycentric coordinates*. Show furthermore that the barycentric coordinates (as functions of x) coincide with the three nodal basis functions for the vertices of T.

PROBLEM 1.18. Start from the example triangulation from Figure 1 and plot the interpolation of the function $u(x, y) = \sin(12\pi x)y^2$ on a sequence of 6 red-refined triangulations.

PROBLEM 1.19. Prove the unproven assertions from Theorem 1.24.

PROBLEM 1.20. Compute the kernel of the local stiffness matrix.

PROBLEM 1.21. Study all the routines of this section line by line and convince yourself that they are doing what they are expected to do.

PROBLEM 1.22. Do a convergence study for the unit square and the right-hand side $f(x) = 2(x_1(1 - x_1) + x_2(1 - x_2))$ (exact solution see above) with respect to the following error (semi-)norm

$$\|\nabla(u - u_h)\|_{L^2(\Omega)}$$

similar to that from the above convergence test. For computing the gradient of u_h on a given element T, use the local representation in terms of the nodal basis. The gradients of the basis vectors were already computed in the loop for the stiffness matrix. Perform an analogous convergence study for the error in the L^2 norm and compare the convergence rates (with respect to the maximal diameter of the triangles in the triangulations, the so-called mesh size). Visualize the results in a loglog-diagram (horizontal axis: mesh size, vertical axis: error in the different norms).

PROBLEM 1.23. Given a triangle T with barycentric coordinates (nodal basis functions) φ_1, φ_2, φ_3, prove the formula

$$\int_T \varphi_1^a \varphi_2^b \varphi_3^c \, dx = 2|T| \frac{a!b!c!}{(a + b + c + 2)!}$$

for any $a, b, c \in \mathbb{N} \cup \{0\}$. (It is enough to show it on the reference triangle with vertices $(0, 0)$, $(1, 0)$, $(0, 1)$ and to then argue by transformation.)

PROBLEM 1.24. Show that the finite element space satisfies $S^1(\mathcal{T}) \subseteq H^1(\Omega)$.

PROBLEM 1.25. Let $\mathcal{T}$ be a regular triangulation of $\Omega \subseteq \mathbb{R}^2$ and let $v \in P_1(\mathcal{T})$ be a piecewise affine function. For each interior edge F with adjacent triangles T_+ and T_- (i.e., $F = T_+ \cap T_-$), the jump across F is defined by $[v]_F := v|_{T_+} - v|_{T_-}$.

(a) Prove that

$$v \in H^1(\Omega) \iff [v]_F = 0 \quad \text{for all interior edges } F.$$

(b) The space $H(\mathrm{div}, \Omega)$ is defined by

$$H(\mathrm{div}, \Omega) := \left\{ v \in L^2(\Omega; \mathbb{R}^2) \;\middle|\; \begin{array}{l} \exists g \in L^2(\Omega) \text{ such that for all } \varphi \in C_c^\infty(\Omega) \\ \int_\Omega v \cdot \nabla \varphi \, dx = -\int_\Omega g \varphi \, dx \end{array} \right\}.$$

Prove that

$$v \in H(\mathrm{div}, \Omega) \iff [v \cdot \nu_F]_F = 0 \quad \text{for all interior edges } F$$

where ν_F is some normal vector of F.

PROBLEM 1.26. Show that $\| \cdot \|_{H^1(\Omega)}$ is a norm on $H^1(\Omega)$. Does $\|\nabla \cdot \|_{L^2(\Omega)}$ define a norm on $H^1(\Omega)$ as well?

PROBLEM 1.27. Let $v(x) = \log(|\log(|x|)|)$ be given on the disc $\Omega = \{x \in \mathbb{R}^2 : |x| < 1/\exp(1)\}$. Prove $v \in H^1(\Omega)$ (cf. Problem 1.8).

PROBLEM 1.28. Prove that $H^1(\Omega)$ is complete with respect to $\| \cdot \|_{H^1(\Omega)}$. (*Hint:* You may use the result that $L^2(\Omega)$ is complete (Fischer-Riesz Theorem)).

PROBLEM 1.29. Let $v, w \in H^1(\Omega)$. Prove that the product vw is weakly differentiable and satisfies $\partial_j(vw) = (\partial_j v)w + v(\partial_j w)$. Does vw belong to $H^1(\Omega)$?

PROBLEM 1.30. Let $\Omega \subset \hat{\Omega}$ be bounded open sets and let $u \in H^1(\hat{\Omega})$ be a function with compact support within $\hat{\Omega}$. Prove that the shifted function $u_t(x) = u(x+tb)$ for some fixed vector $b \in \mathbb{R}^2$ satisfies $\|u_t - u\|_{H^1(\Omega)} \to 0$ for $t \to 0$. *(Hint: Approximate u in the L^2 norm by a smooth function ϕ and use uniform continuity of ϕ.)*

PROBLEM 1.31. Let T be a triangle with with set of vertices $\mathcal{N}(T)$. Given $y \in \mathcal{N}(T)$, denote by $\varphi_y \in P_1(T)$ local hat function with

$$\varphi_y(z) = \delta_{yz} \quad \text{for all } z \in \mathcal{N}(T).$$

Compute the following 3×3 matrices

$$M_T := \left(\int_T \varphi_y \varphi_z \, dx \right)_{(y,z) \in (\mathcal{N}(T))^2} \qquad \text{(local mass matrix)}$$

$$C_T := \left(\int_T \varphi_y (\beta \cdot \nabla \varphi_z) \, dx \right)_{(y,z) \in (\mathcal{N}(T))^2} \qquad \text{(local convection matrix; } \beta \in \mathbb{R}^2\text{)}$$

$$S_T := \left(\int_T \nabla \varphi_y \cdot \nabla \varphi_z \, dx \right)_{(y,z) \in (\mathcal{N}(T))^2} \qquad \text{(local stiffness matrix).}$$

You may use the formula from Problem 1.23. The gradients can be assumed to be given as a matrix $[\nabla \varphi_j^\top]_{j=1}^3$ as in prior sections.

PROBLEM 1.32. Show that any function $u \in C^1(\bar{\Omega}) \cap C^2(\Omega)$, satisfying $-\Delta u = f$ for $f \in C^0(\bar{\Omega})$ and $u|_{\partial\Omega} = 0$, also satisfies the variational formulation.

PROBLEM 1.33. Let $T \subseteq \mathbb{R}^2$ be a triangle and $v \in H^2(T) := \{w \in H^1(T) : \partial_j w \in H^1(T) \text{ for } j = 1, 2\}$ with norm

$$\|v\|_{H^2(T)} = \sqrt{\sum_{|\alpha| \le 2} \|\partial^\alpha v\|^2_{L^2(T)}}.$$

(a) Consider a sub-triangle $t := \operatorname{conv}\{A, B, C\}$ with $E := \operatorname{conv}\{A, B\}$ and with tangent vector τ. Apply the trace inequality to $f|_E := \nabla v \cdot \tau$ and prove that

$$|v(B) - v(A)| \le |E|^{1/2} \varrho^{-1/2} 2\big(1 + \operatorname{diam}(t)^2\big)^{1/2} \|v\|_{H^2(t)}$$

for $\varrho := 2|t|/|E|$.

(b) For any two points A and B in T there exists $C \in T$ such that (with $E := \operatorname{conv}\{A, B\}$ and $t := \operatorname{conv}\{A, B, C\}$), ϱ^{-1} is uniformly bounded by some constant $C(T)$ that depends only on T, but not on A, B, or t.

(c) Conclude that v is Hölder continuous with exponent $1/2$.

Remark: This shows the embedding $H^2(T) \hookrightarrow C^{0,1/2}(T)$ on a triangle.

PROBLEM 1.34. (a) Prove that in one space dimension the approximation of the equation $u''(x) = 1$ (on the interval $(0, 1)$ with homogeneous Dirichlet boundary conditions) with the P_1 finite element method results in the nodal interpolation, that is $u_h = I_h u$.

(b) Convince yourself that this property cannot be valid in higher space dimensions (e.g., by a computational test case).

PROBLEM 1.35. Let $f \in L^2(\Omega)$ and recall the energy functional

$$J(v) := \frac{1}{2}\|\nabla v\|^2_{L^2(\Omega)} - \int_\Omega f v \, dx \quad \text{for } v \in H^1_0(\Omega).$$

Prove that the error of the finite element method for the Poisson problem with right-hand side f satisfies

$$\|\nabla(u - u_h)\|^2_{L^2(\Omega)} = 2(J(u_h) - J(u)) = \|\nabla u\|^2_{L^2(\Omega)} - \|\nabla u_h\|^2_{L^2(\Omega)}.$$

PROBLEM 1.36. (a) Write the data structures for a triangulation of the L-shaped domain $\Omega := (-1, 1)^2 \setminus ([0, 1] \times [-1, 0])$ with Dirichlet boundary $\partial\Omega$.

(b) Plot the convergence history for $-\Delta u = 1$ on the L-shaped domain (cf. Problem 1.35; the exact solution satisfies $\|\nabla u\|^2 = 0.2140750232$). Compare the convergence rate with the results on the square domain.

PROBLEM 1.37. Prove the assertions of Theorem 1.46 and track the dependence of the constant C on the spectral bounds a_0, a_1 and the L^∞ norms of the coefficients.

PROBLEM 1.38. Prove existence and uniqueness of solutions to the second-order elliptic problem in case of nontrivial Neumann boundary $\Gamma_N \neq 0$.

PROBLEM 1.39. *(convection-diffusion equation)*

(a) Implement the finite element method for the convection-diffusion equation $-\varepsilon\Delta u + \beta \cdot \nabla u = f$.

(b) Consider the unit square $\Omega = (0, 1)^2$ with homogeneous Dirichlet boundary conditions and the right-hand side f according to the exact solution

$$u(x) = \left(\frac{e^{r_1(x_1-1)} - e^{r_2(x_1-1)}}{e^{-r_1} - e^{-r_2}} + x_1 - 1 \right) \sin(\pi x_2)$$

with

$$r_1 = \frac{-1 + \sqrt{1 + 4\varepsilon^2\pi^2}}{-2\varepsilon} \quad \text{and} \quad r_2 = \frac{-1 - \sqrt{1 + 4\varepsilon^2\pi^2}}{-2\varepsilon}.$$

Run numerical computations for the following parameters

(i) $\varepsilon = 0.1$ and $\beta = (1, 0)^T$.

(ii) $\varepsilon = 0.001$ and $\beta = (1, 0)^T$.

PROBLEM 1.40. Extend your finite element code to the case of inhomogeneous Dirichlet and zero Neumann boundary data. Use the follwing test case to validate your code (via comparison of the solution graphs and convergence tests): The domain $\Omega = (0,1)^2$ is the unit square. The Neumann boundary is the line $\{(1,t) : 0 < t < 1\}$, and $\Gamma_D = \partial\Omega \setminus \Gamma_N$. The exact solution is given by

$$u(x,y) = 5 + \sin(\frac{\pi}{2}x)\sin(\pi y)$$

with $u_D = 5$ and $g = 0$. The right-hand side reads $f = \frac{5}{4}\pi^2 \sin(\frac{\pi}{2}x)\sin(\pi y)$.

PROBLEM 1.41. *(nodal interpolation not L^2 or H^1 stable)* For a triangle $T \subseteq \mathbb{R}^2$, prove that there is no constant C such that the nodal P_1 interpolation I satisfies

$$\|Iu\|_{L^2(T)} \le C\|u\|_{L^2(T)} \text{ for all } u \in C^\infty(T)$$
$$\text{or} \quad \|\nabla Iu\|_{L^2(T)} \le C\|\nabla u\|_{L^2(T)} \text{ for all } u \in C^\infty(T).$$

PROBLEM 1.42. Prove Theorem 1.52.

PROBLEM 1.43. Let $\mathcal{T}$ be a triangulation. Prove that the aspect ratio of the triangles stays bounded under iterative red refinement.

PROBLEM 1.44. Prove that there exists a constant C only dependent on the shape regularity such that any finite element function $v_h \in S^1(\mathcal{T})$ satisfies

$$\|\nabla v_h\|_{L^2(T)} \le Ch_T^{-1}\|v_h\|_{L^2(T)} \quad \text{for all } T \in \mathcal{T}.$$

This estimate is called *inverse inequality. (Hint: Use transformation to a reference element $\hat{T}$. Use equivalence-of-norms argument in the finite dimensional space $P_1(\hat{T})$ with a constant $C(\hat{T})$ only depending on $\hat{T}$. Afterwards, transform back.)*

PROBLEM 1.45. Prove for the solution u from Example 1.13 that $u \notin H^2(\Omega)$.

PROBLEM 1.46. A family of triangulations satisfies the *minimal angle condition* if there is a lower bound $0 < \alpha_0$ to all interior angles of the triangles from that family. Prove that the *minimal angle condition* implies shape regularity.

PROBLEM 1.47. Prove that the nodal basis functions φ_z on a triangle T satisfy

$$\|\nabla\varphi_z\|_{L^2(T)} \le C_1 \quad \text{and} \quad \|\nabla\varphi_z\|_{L^\infty(T)} \le C_2 h_T^{-1}$$

with constants C_1, C_2 only depending on the shape regularity.

PROBLEM 1.48. Prove that the L^2 inner product on finite element spaces is represented by the *mass matrix*

$$M_{jk} = \int_\Omega \varphi_j\varphi_k \, dx.$$

Prove that the local mass matrix is given by

$$M_{jk} = \frac{|T|}{12}\begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

(see also Problem 1.23). Implement an assembling routine for M in Python.

PROBLEM 1.49. Prove the approximation properties of the difference quotients stated in Lemma 1.70.

PROBLEM 1.50.
- Implement the backward and the forward Euler method for the heat equation. Approximate g by the finite element interpolation $g_h = I_h g$.

- Let the initial value $u_0(x) = \sin(\pi x_1)\sin(\pi x_2)$ on the unit square $\Omega = (0,1)^2$ be given. Prove that the solution to the heat equation with $f = 0$ is given by
$$u(t,x) = \sin(\pi x_1)\sin(\pi x_2)\exp(-2\pi^2 t).$$
- Take these data (and $T = 1$) and compute experimental convergence rates with respect to h and Δt. Use the following two choices for the norm:
$$\max_{k=0,\dots,N} \|\nabla u - \nabla u_h^k\|_{L^2(\Omega)}$$
and
$$\sqrt{\int_\Omega \|\nabla u(t) - \nabla u_h(t)\|^2_{L^2(\Omega)}\,dt}.$$
The last integral can be approximated by the midpoint rule in space and the Simpson rule in time.

  Remark: We interpret $u_h(t)$ to be piecewise affine in time (a polygonal line through the points u_h^k).

PROBLEM 1.51. Prove the error estimate
$$\max_{k=1,\dots,J} \|u(\cdot,t_k) - u_h^k\|_{L^2(\Omega)} \le \mathcal{O}(h^2 + \Delta t).$$
from Theorem 1.73.

PROBLEM 1.52 (H^1 stability of the L^2 projection). Let $\Pi_h : H^1_0(\Omega) \to S^1_0(\mathcal{T})$ denote the L^2 projection, i.e., for given $v \in H^1_0(\Omega)$, the function $\Pi_h v \in S^1_0(\mathcal{T})$ satisfies
$$\int_\Omega (\Pi_h v) w_h\,dx = \int_\Omega v w_h\,dx \quad \text{for all } w_h \in S^1_0(\mathcal{T}).$$
Then, clearly, $\|\Pi_h v\|_{L^2(\Omega)} \le \|v\|_{L^2(\Omega)}$. Prove that for a family of red-refined triangulations there exists a constant C such that
$$\|\nabla \Pi_h v\|_{L^2(\Omega)} \le C\|\nabla v\|_{L^2(\Omega)},$$
i.e., the L^2 projection is H^1 stable.
Hint: Given v, prove $\Pi_h v = \Pi_h(R_h v - v) + R_h v$. For the first term, use an inverse estimate (Problem 1.44), the L^2 stability of the projection $\Pi_h v$, and the approximation and stability properties of the quasi-interpolation R_h.

PROBLEM 1.53 (density of finite element spaces). Prove that the finite element spaces $S^1_0(\mathcal{T}_j)$ with respect to a sequence $\mathcal{T}_j$ of red refined triangulations are dense in $H^1_0(\Omega)$.
Hint: Given v, approximate it by a smooth function v_ε and interpolate v_ε by a finite element function on a sufficiently fine mesh.

CHAPTER 2

Advanced Finite Element Methods

2.1. Galerkin method

2.1.1. Closed range theorem and Banach–Babuška–Nečas lemma. We want to characterize isomorphisms between certain Banach spaces. We start by recalling a version of the Hahn–Banach theorem from linear functional analysis.

THEOREM 2.1 (Hahn–Banach). *Let $M \subseteq X$ be a subspace of the normed linear space $(X, \|\cdot\|_X)$ and let $f \in M^*$. Then there exists $F \in X^*$ such that $F|_M = f$ and $\|F\|_{X^*} = \|f\|_{M^*}$.*

PROOF. This is taught in any class on linear functional analysis. □

As a consequence, we note the following fundamental separation property.

THEOREM 2.2 (separation). *Let $M \subseteq X$ be a closed subspace of the Banach space X and let $z \in X \setminus M$ be a point outside M. Then there exists $F \in X^*$ with $\|F\|_{X^*} = 1$ that satisfies $F|_M = 0$ and $F(z) = \operatorname{dist}(z, M)$.*

PROOF. We construct the linear functional f on $\tilde{M} = M + \operatorname{span}\{z\}$ by

$$f(y + \alpha z) = \alpha \operatorname{dist}(z, M) \quad \text{for any } y \in M, \alpha \in \mathbb{R}.$$

We compute

$$|f(y + \alpha z)| \leq |\alpha| \operatorname{dist}(z, M) \leq |\alpha| \, \|z + \alpha^{-1} y\|_X = \|\alpha z + y\|_X$$

which shows continuity of f, that is, $f \in \tilde{M}^*$ and $\|f\|_{\tilde{M}^*} \leq 1$. By the definition of the distance and the closedness of M, we have that, given any $\varepsilon > 0$, there exists $y_\varepsilon \in M$ such that $\|z - y_\varepsilon\|_X \leq (1 + \varepsilon) \operatorname{dist}(z, M)$ such that $f(z - y_\varepsilon) \geq (1 + \varepsilon)^{-1} \|z - y_\varepsilon\|_X$. Thus, $\|f\|_{\tilde{M}^*} \geq 1$. We now apply the Hahn–Banach theorem to $\tilde{M}$ and f, which shows the existence of the claimed extension F. □

We use the notation $\langle f, v \rangle = f(v)$.

DEFINITION 2.3 (annihilator, polar set). Let X be a Banach space with a subspace $V \subseteq X$ and let $U \subseteq X^*$ be a subspace of its dual. We define the *annihilator* of V by

$$V^\circ := \{f \in X^* : \langle f, v \rangle = 0 \text{ for all } v \in V\} \subseteq X^*$$

and the *polar set* of U by

$$^\circ U := \{x \in X : \langle u, x \rangle = 0 \text{ for all } u \in U\} \subseteq X.$$

♦

We have the following elementary property.

LEMMA 2.4 (characterization of the closure). *Let $V \subseteq X$ be a subspace of a Banach space X. Then $^\circ(V^\circ) = \overline{V}$.*

PROOF. The space $^\circ(V^\circ)$ is the intersection of kernels of continuous linear operators and is therefore closed. The definitions imply that any $x \in V$ satisfies $x \in {}^\circ(V^\circ)$. Since $^\circ(V^\circ)$ is closed we therefore have $\overline{V} \subseteq {}^\circ(V^\circ)$. By the separation theorem, any $z \notin \overline{V}$ can be separated from $^\circ(V^\circ)$, i.e., there exists $F \in V^\circ$ with $F(z) \neq 0$, whence $z \notin {}^\circ(V^\circ)$. This shows the claimed equality of spaces. □

We recall that for Banach spaces X and Y and a continuous linear map $L : X \to Y$ the dual $L^* : Y^* \to X^*$ is defined by

$$L^*(F) = \langle F, L\,\cdot\rangle \in X^*.$$

We recall the closed range theorem. We denote by $\mathcal{L}(X, Y)$ the space of bounded and continuous maps from X to Y.

THEOREM 2.5 (closed range theorem). *Let $L \in \mathcal{L}(X, Y)$ be a continuous linear map between Banach spaces X and Y. The range $L(X)$ is closed in Y if and only if $L(X) = {}^\circ(\ker L^*)$.*

PROOF. We have $f \in \ker L^*$ if and only if $\langle f, Lx\rangle = 0$ for all $x \in X$, which means $f \in L(X)^\circ$. We apply the foregoing lemma to the space $\ker L^* = L(X)^\circ$ and conclude the proof. □

The main application of the closed range theorem for our purposes is the characterization of solvability of operator equations. Recall that a Banach space Y is called reflexive if the map

$$J : Y \to Y^{**}, \quad Y \ni y \mapsto \langle \cdot, y\rangle$$

from Y to its bidual Y^{**} is an isomorphism.

LEMMA 2.6 (Banach–Babuška–Nečas lemma). *Let X be a Banach space and let Y be a reflexive Banach space. A linear map $L : X \to Y^*$ is an isomorphism if and only if the following three conditions are satisfied:*

1. *Continuity: $\|Lx\|_{Y^*} \leq C\|x\|_X$ for a constant $C > 0$ and all $x \in X$.*
2. *There exists $\gamma > 0$ such that for all $x \in X$*
$$\gamma\|x\|_X \leq \|Lx\|_{Y^*}.$$
3. *For every nonzero $y \in Y \setminus \{0\}$ there exists some $x \in X$ such that $\langle Lx, y\rangle \neq 0$.*

PROOF. Let conditions (1)–(3) be satisfied. Then, by (1), L is continuous and, by (2), it is injective because $Lx = 0$ implies $x = 0$. Hence, L is bijective as a map from X to its range $L(X)$. The inverse $L^{-1} : L(X) \to X$ is continuous because, by (2),

$$\|L^{-1}z\|_X \leq \gamma^{-1}\|LL^{-1}z\|_{Y^*} = \gamma^{-1}\|z\|_{Y^*}.$$

The continuity of L and L^{-1} implies that $L(X)$ is closed. The closed range theorem then teaches

$$L(X) = {}^\circ(\ker L^*) \subseteq Y^*. \tag{20}$$

Let us write down the polar set of $\ker L^* \subseteq Y^{**}$ explicitly:

$${}^\circ(\ker L^*) = \{v \in Y^* : \langle u, v\rangle = 0 \text{ for all } u \in \ker L^*\}.$$

We furthermore observe from the definition of L^* that

$$u \in \ker L^* \iff \langle L^*u, x\rangle = 0 \text{ for all } x \in X \iff \langle u, Lx\rangle = 0 \text{ for all } x \in X.$$

Since Y is reflexive, we see that

$$u \in \ker L^* \iff \langle Lx, J^{-1}u\rangle = 0 \text{ for all } x \in X$$

for $J^{-1}u \in Y$ and the isomorphism J. Property (3) therefore implies that $J^{-1}u = 0$ and so $\ker L^* = \{0\}$. By (20), we then have that $L(X) = Y^*$. Thus, L is an isomorphism.
The proof of the converse direction is immediate and left as an exercise to the readers. □

Condition (2) of Lemma 2.6 is often called the inf-sup condition because γ can be represented as

$$\gamma = \inf_{x\in X\setminus\{0\}} \sup_{y\in Y\setminus\{0\}} \frac{\langle Lx, y\rangle}{\|x\|_X \|y\|_Y}.$$

2.1.2. Quasi-optimality of the Galerkin method. We consider the situation of a Banach space X and reflexive Banach space Y^*. Suppose and $x \in X$ and $f \in Y^*$ satisfy $Lx = f$. In any practical simulation we need to approximate the infinite-dimensional spaces X and Y. Suppose we are given closed subspaces $X_h \subseteq X$ and $Y_h \subseteq Y$ with the inclusion mappings ι_X and ι_Y. Then the *Galerkin method* is to find $x_h \in X_h$ such that Lx_h equals f when restricted to test functions of Y_h.

THEOREM 2.7 (Galerkin method). *Consider a Banach space X and reflexive Banach Y with closed subspaces $X_h \subseteq X$ and $Y_h \subseteq Y$ with $L \in \mathcal{L}(X, Y^*)$ and let $x \in X$ solve $Lx = f$ for some $f \in Y^*$. Assume that there exists $\gamma_h > 0$ such that*

$$\gamma_h \le \inf_{\xi_h\in X_h\setminus\{0\}} \frac{\|L\xi_h\|_{Y_h^*}}{\|\xi_h\|_X}$$

and that for any $y_h \in Y_h \setminus \{0\}$ there exists $\xi_h \in X_h$ with $\langle L\xi_h, y_h\rangle \neq 0$. Then there exists a unique solution $x_h \in X_h$ to $\iota_Y^* L x_h = \iota_Y^* f$. It satisfies the error bound

$$\|x - x_h\|_X \le (1 + \frac{\|\iota_Y^* L\|_{\mathcal{L}(X,Y^*)}}{\gamma_h}) \inf_{z_h\in X_h} \|x - z_h\|_X.$$

PROOF. The existence and uniqueness of x_h follow from the Banach–Babuška–Nečas lemma. For any $z_h \in X_h$ we have that

$$\gamma_h \|z_h - x_h\|_X \le \|L(z_h - x_h)\|_{Y_h^*} = \sup_{y_h\in Y_h\setminus\{0\}} \frac{\langle L(z_h - x_h), y_h\rangle}{\|y_h\|_Y}. \tag{21}$$

Since $\langle Lx_h, y_h\rangle = \langle f, y_h\rangle$ from the solution property of x_h, we deduce from the continuity of L that

$$\gamma_h \|z_h - x_h\|_X \le \|\iota_Y^* L\|_{\mathcal{L}(X,Y^*)} \|\langle L(z_h - x), y_h\rangle\|.$$

The asserted bound follows from the triangle inequality $\|x - x_h\|_X \le \|x - z_h\|_X + \|z_h - x_h\|_X$ and the infimum over z_h. □

The main application is that X_h and Y_h are finite-dimensional. Condition (21) is then called the discrete inf-sup condition. The nondegeneracy assumption means that the spaces have the same dimension. In practice, we think of h being a mesh parameter that increases the resolution by being decreased. The error bound for the Galerkin method is proportional to γ_h^{-1} and, therefore, it is important to have the inf-sup condition *uniformly in h*.

EXAMPLE 2.8. In variationally formulated and coercive PDEs over a Hilbert space X we choose $X = Y$ and a bilinear form $a : X \times X \to \mathbb{R}$. It induces a linear operator $L : X \to X^*$ by $x \mapsto a(x, \cdot)$. Given $f \in X^*$, the equation $Lx = f$ then means

$$a(x, y) = \langle f, y\rangle \quad \text{for all } y \in X.$$

For $X_h \subseteq X$ as above, the discrete equation $\iota_Y^* L x_h = \iota_X^* f$ means

$$a(x_h, y_h) = \langle f, y_h \rangle \quad \text{for all } y_h \in X_h$$

and should be familiar to the reader from previous elementary courses. As the most important example we mention $X = H_0^1(\Omega)$ as the usual Sobolev space over a suitable domain Ω and the form a related to an elliptic second-order operator. The above theorem then resembles Céa's lemma.

2.1.3. Saddle-point problems in reflexive spaces. Minimization of functionals subject to linear constraints can be re-formulated with Lagrange multipliers. The usual (formal) derivation of a necessary condition of

$$\text{minimize } \frac{1}{2}\langle Au, u\rangle - \langle f, u\rangle \quad \text{over } V \text{ subject to } Bu = 0$$

is to introduce a Lagrange multiplier $p \in M$ such that

$$Au + B'u = f$$

with an adjoint operator $B' := B^* J_M$. We want to study the well-posedness of such formulations. In this situation, we are given a product spaces $X = V \times M$ where the operator L has block structure. Given $F \in V^*$ and $G \in M^*$, a so-called *saddle-point problem* has the format

$$L \begin{bmatrix} u \\ p \end{bmatrix} := \begin{bmatrix} A & B' \\ B & 0 \end{bmatrix} \begin{bmatrix} u \\ p \end{bmatrix} = \begin{bmatrix} F \\ G \end{bmatrix}. \tag{22}$$

The conditions of the Banach–Babuška–Nečas lemma can equivalently formulated as conditions on A and B. We are given a bounded linear B operator that is not surjective but has a bounded inverse on its range. In finite dimensions, these are the full rank rectangular matrices. We want to study analogous mapping properties in reflexive Banach spaces.

LEMMA 2.9. *Let V and M be reflexive Banach spaces and $B \in \mathcal{L}(V, M^*)$, with $Z := \ker B$, satisfy*

$$0 < \beta = \inf_{\mu \in M \setminus \{0\}} \frac{\|B^* J_M \mu\|_{V^*}}{\|\mu\|_M}. \tag{23}$$

Then, $B^* J_M : M \to Z^\circ$ is an isomorphism with $\|(B^* J_M)^{-1}\|_{\mathcal{L}(Z^\circ, M)} \leq \beta^{-1}$.

PROOF. We observe that the range of $B^* J_M$ is indeed a subset of Z° because $\langle B^* J_M \mu, z\rangle = \langle J_M \mu, Az\rangle = 0$ for any $\mu \in M$ and any $z \in Z$. By the above assumptions, $B^* J_M$ is continuous (property (1) of Lemma 2.6) and (23) implies that property (2) of Lemma 2.6 is satisfied. As in the proof of Lemma 2.6 we therefore see that $B^* J_M$ and its inverse are continuous. The closed range theorem then shows

$$B^* J_M(M) = {}^\circ(\ker((B^* J_M)^*)).$$

It is direct to verify

$$u \in \ker((B^* J_M)^*) \iff J_V^{-1} u \in Z.$$

Thus the range equals Z° and we have established the isomorphism. The bound on the norm is left as an exercise. □

Brezzi's splitting theorem performs block elimination in the above system. It is our main criterion for saddle-point problems.

THEOREM 2.10 (Brezzi splitting). *Let V and M be reflexive Banach spaces and $B \in \mathcal{L}(V, M^*)$, with $Z := \ker B$. The operator*

$$L : X \to X^* \quad \text{with } X = V \times M$$

from (22) *is an isomorphism if and only if A is an isomorphism from Z to Z^* with*

$$0 < \alpha = \inf_{z \in Z \setminus \{0\}} \frac{\|Az\|_{Z^*}}{\|z\|_V}$$

and B satisfies the inf-sup condition (23)*. Given $F \in V^*$ and $G \in M^*$, the unique solution $(u, p) \in V \times M$ to* (22) *block satisfies*

$$\|u\|_V \leq \alpha^{-1}\|F\|_{V^*} + \beta^{-1}(1 + \frac{C_A}{\alpha})\|G\|_{M^*},$$
$$\|p\|_M \leq \beta^{-1}(1 + \frac{C_A}{\alpha})\|F\|_{V^*} + \beta^{-1}(1 + \frac{C_A}{\alpha})\frac{C_A}{\beta}\|G\|_{M^*}.$$

Here, $C_A = \|A\|_{\mathcal{L}(V,V^*)}$.

Proof. We have seen in the previous lemma that $B^*J_M : M \to Z^\circ$ is an isomorphism, and so is $(B^*J_M)^* : (Z^\circ)^* \to M^*$ with the same continuity constant for the inverse. Hence, for the given $G \in M^*$ there exists $\eta \in (Z^\circ)^*$ with $(B^*J_M)^*\eta = G$ with $\|\eta\|_{(Z^\circ)^*} \leq \beta^{-1}\|G\|_{M^*}$. We denote by $\hat{\eta} \in V^{**}$ the Hahn–Banach extension of η that coincides with η on Z° and has the same norm. Then, the element $u_0 := J_V^{-1}\hat{\eta}$ satisfies for any $\mu \in M$ that

$$\begin{aligned}\langle Bu_0, \mu\rangle = \langle J_M\mu, Bu_0\rangle = \langle B^*J_M\mu, u_0\rangle = \langle B^*J_M\mu, J_V^{-1}\hat{\eta}\rangle \\ = \langle \hat{\eta}, B^*J_M\mu\rangle = \langle (B^*J_M)^*\eta, \mu\rangle = \langle G, \mu\rangle.\end{aligned}$$

Hence, $Bu_0 = G$ with $\|u_0\|_V \leq \beta^{-1}\|G\|_{M^*}$. Upon defining $w := u - u_0$, we reformulate the original problem into

$$Aw + B'p = F - Au_0$$
$$Bw = 0.$$

We restrict the first equation to Z and obtain from the assumed isomorphism property of A that there exists a unique $w \in Z$ satisfying

$$\iota_Z^* Aw = \iota_Z^*(F - Au_0)$$

with $\|w\|_V \leq \alpha^{-1}(\|F\|_{V^*} + C_a\|u_0\|_V)$. Here, ι_Z is the inclusion of Z to V (this notation is short, but not consistent with the above one).
Since $F - A(u_0 + w) \in Z^\circ$, the foregoing lemma yields the existence of $p \in M$ with

$$B'p = F - A(u_0 + w)$$

and

$$\|p\|_M \leq \beta^{-1}(\|F\|_{V^*} + C_a\|u_0 + w\|_V).$$

Hence, $u := u_0 + w$ and p solve the saddle-point problem. The asserted norm bounds follow from directly tracing the constants in the above estimates. The proof of the converse statement is left as an exercise. □

Remark 2.11. The saddle-point problem is encountered more often in a variational form in the literature and reads as

$$\begin{aligned} a(u, v) + b(v, p) &= F(v) \quad \text{for all } v \in V \\ b(u, q) &= G(q) \quad \text{for all } q \in M \end{aligned} \tag{24}$$

for bounded bilinear forms $a : V \times V \to \mathbb{R}$ and $b : V \times M \to \mathbb{R}$. This is of course equivalent to the above formulation. Indeed, we see that $Au := a(u, \cdot) \in V^*$ and $Bu := b(u, \cdot) \in M^*$. It is easy to check that $B'\mu$ then equals $b(\cdot, \mu) \in V^*$ and that the kernel can be written as

$$Z = \{v \in V : \forall \mu \in M b(v, \mu) = 0\}.$$

♦

If we want to discretize the saddle-point problem with closed subspaces $V_h \subseteq V$ and $M_h \in\subseteq M$, we can derive well-posedness and an error bound by applying Theorem 2.7 to $X_h = V_h \times M_h$. We apply Theorem 2.10 to the discrete setting and see that the (global) discrete inf-sup condition follows from the conditions

$$0 < \alpha_h = \inf_{z_h \in Z_h \setminus \{0\}} \frac{\|Az_h\|_{Z_h^*}}{\|z_h\|_V} \quad \text{and} \quad 0 < \beta_h = \inf_{\mu_h \in M_h \setminus \{0\}} \frac{\|B'\mu_h\|_{V_h^*}}{\|\mu\|_{M_h}}. \tag{25}$$

Here, $Z_h := \ker(\iota_M^* B \iota_V)$, equivalently written as

$$Z_h = \{v_h \in V_h : \forall \mu_h \in M_h \ \langle Bv_h, \mu_h \rangle = 0\}$$

is the *discrete kernel*. It is **very important** to note that in general we must expect $Z_h \not\subseteq Z$. Usually, the condition on α_h is not very critical, for example when A is coercive. The condition on β_h is very delicate and is linked to the compatibility of the two discrete spaces. We note the following consequence.

Corollary 2.12. *Let the conditions of Theorem 2.10 hold and let the closed subspaces $V_h \subseteq V$ and $M_h \subseteq M$ satisfy (25). Let $(u, p) \in V \times M$ solve the saddle-point problem with right-hand side (F, G). Then, there exists a unique pair $(u_h, p_h) \in V_h \times M_h$ solving*

$$\begin{aligned} \iota_V^* A u_h + \iota_V^* B' p_h &= \iota_V^* F \\ \iota_M^* B p_h &= \iota_M^* G. \end{aligned}$$

It satisfies

$$\|u - u_h\|_V + \|p - p_h\|_M \leq C\Big(\inf_{v_h \in V_h} \|u - v_h\|_V + \inf_{q_h \in M_h} \|p - q_h\|_M\Big)$$

with a constant C that only depends on α_h, β_h, C_A.

2.2. Stokes equations

2.2.1. The Stokes equation. We now draw our attention to problems under linear constraints. For example, the velocity field $u : \Omega \to \mathbb{R}^2$ of a very viscous fluid under some volume force $f : \Omega \to \mathbb{R}^2$ is modelled by the following constrained minimization problem:

$$J(v) := \frac{1}{2} \int_\Omega |Dv|^2 \, dx - \int_\Omega f \cdot v \, dx \to \min \quad \text{subject to } \operatorname{div} u = 0 \quad \text{and } u|_{\partial\Omega} = 0.$$

The energy is measured with the Dirichlet functional for vector-valued variables. The constraint $\operatorname{div} u = 0$ models that the fluid under consideration is incompressible. In order to —at least formally— compute a corresponding PDE as an Euler–Lagrange equation, we need to reformulate it as a problem on the whole nonconstrained space. The enegry J is well-defined on $[H_0^1(\Omega)]^2$, the space of vector fields whose components belong to $H_0^1(\Omega)$. This can be done via Lagrange multipliers. The idea is to add a term involving the constraint to the functional and to minimize

$$\frac{1}{2} \int_\Omega |Dv|^2 \, dx - \int_\Omega p \operatorname{div} v - \int_\Omega f \cdot v \, dx$$

for all $v \in [H_0^1(\Omega)]^2$ and some p in the range of the divergence operator. The latter is then called the space of Lagrange multipliers. It will turn out that the range is precisely

$$L_0^2(\Omega) = \{v \in L^2(\Omega) : \int_\Omega v \, dx = 0\}.$$

As in the proof of the Dirichlet principle (Theorem 1.15), we can compute the derivatives in the directions of perturbations to v and to q and arrive at the following necessary condition

$$\begin{aligned}\int_\Omega Du : Dv\,dx - \int_\Omega p \operatorname{div} v\,dx &= \int_\Omega f\cdot v\,dx \qquad \text{for all } v\in [H^1_0(\Omega)]^2,\\ -\int_\Omega q\operatorname{div} u\,dx &= 0 \qquad\qquad\qquad \text{for all } q\in L^2_0(\Omega)\end{aligned}$$

for the solution pair $(u,p)\in [H^1_0(\Omega)]^2\times L^2_0(\Omega)$. (The notation $A:B=\sum_{j=1}^2 A_{jk}B_{jk}$ is used the inner product of matrices.) The physical interpretation of the Lagrange multiplier p is the role of a pressure variable. This system is referred to as the *Stokes equations*. It is easy to see that this system is symmetric but not coercive with respect to the product space (chose for instance $p=\lambda\operatorname{div} u$ for sufficiently large λ). Recall that systems of this structure are called *saddle-point* problems.
We will use the Brezzi splitting theorem to prove well-posedness of the Stokes equations. Obviously, the Stokes equations are equivalent to (24) with the choices $V=[H^1_0(\Omega)]^2$, $M:=L^2_0(\Omega)$ and

$$a(v,w)=\int_\Omega Dv:Dw\,dx,\quad b(v,q)=-\int_\Omega q\operatorname{div} v\,dx,\quad F(v)=\int_\Omega f\cdot v\,dx,\ G=0.$$

THEOREM 2.13. *Let $\Omega\subseteq\mathbb{R}^2$ be an open, bounded, and connected domain with polygonal Lipschitz boundary. Given any $f\in L^2(\Omega)$, there exists a unique solution $(u,p)\in V\times M$ to the Stokes equations.*

PROOF. Since a is coercive (by Friedrichs' inequality), it remains to prove the inf-sup condition

$$0<\beta=\inf_{q\in L^2_0(\Omega)\setminus\{0\}}\sup_{v\in[H^1_0(\Omega)]^2\setminus\{0\}}\frac{\int_\Omega q\operatorname{div} v\,dx}{\|Dv\|_{L^2(\Omega)}\|q\|_{L^2(\Omega)}}$$

for some β. This result is not shown here because its proof is far beyond the scope of this lecture. It can be found in the literature. □

2.2.2. A finite element method for the Stokes system. For the approximation of saddle-point problems, we aim at choosing finite element subspaces $V_h\subseteq V$ and $M_h\subseteq M$. Since these are closed subspaces, they are again Hilbert spaces and the Brezzi splitting can be used to study the solvability of the resulting discrete problem. In contrast to the coercivity in the Lax–Milgram setting, however, the inf-sup condition is usually not inherited by the discrete spaces. It needs to be imposed as an additional condition. We call

$$0<\beta_h=\inf_{q_h\in M_h\setminus\{0\}}\sup_{v_h\in V_h\setminus\{0\}}\frac{b(v_h,q_h)}{\|v_h\|_V\|q_h\|_M}$$

the *discrete inf-sup condition*. The construction of discrete spaces satisfying this property turns out nontrivial. The standard finite element space $[S^1_0(\mathcal{T})]^2$ is not suited for a discretization involving the constraint on the divergence, see Problem 2.11 and Problem 2.12. We focus for the moment on a practical discretization of the Stokes equations, the so-called Mini finite element. We will define the corresponding spaces, comment on the implementation, and show that it satisfies a discrete inf-sup condition.
Given a triangulation $\mathcal{T}$ of the domain Ω and any $T\in\mathcal{T}$ with vertices z_1, z_2, z_3, we define the so-called element bubble b_T by

$$b_T=\varphi_{z_1}\varphi_{z_2}\varphi_{z_3}$$

where the φ_{z_j} are the hat functions associated to the vertices of T. It is immediate to verify that

- $b_T|_T$ is a cubic polynomial on T,
- b_T is positive in the interior of T,
- b_T vanishes on $\Omega \setminus T$,
- $b_T \in H_0^1(\Omega)$.

We denote the space of bubble functions by

$$\mathcal{B}(\mathcal{T}) := \operatorname{span}\{b_T : T \in \mathcal{T}\}.$$

We then have $\dim(\mathcal{T}) = \#\mathcal{T}$ where $\#$ denotes the cardinality of a set. The Mini finite element discretization is based on the discrete spaces

$$V_h := [S_0^1(\mathcal{T})]^2 \oplus [\mathcal{B}(\mathcal{T})]^2 \quad \text{and} \quad M_h := S^1(\mathcal{T}) \cap L_0^2(\Omega).$$

We clearly have the inclusions $V_h \subseteq V$ and $M_h \subseteq M$. For a practical implementation, we need (local) matrix representations of the bilinear forms a and b.

Local matrices of the Mini finite element. Denote by $\varphi_1, \varphi_2, \varphi_3$ the three nodal basis functions on a triangle T and recall the cubic bubble function $b_T := \varphi_1\varphi_2\varphi_3$. Define the local basis functions of the velocity part of the Mini finite element by

$$\psi_1 = \begin{pmatrix}\varphi_1\\0\end{pmatrix}, \psi_2 = \begin{pmatrix}\varphi_2\\0\end{pmatrix}, \psi_3 = \begin{pmatrix}\varphi_3\\0\end{pmatrix}, \psi_4 = \begin{pmatrix}0\\\varphi_1\end{pmatrix}, \psi_5 = \begin{pmatrix}0\\\varphi_2\end{pmatrix}, \psi_6 = \begin{pmatrix}0\\\varphi_3\end{pmatrix},$$
$$\psi_7 = \begin{pmatrix}b_T\\0\end{pmatrix}, \psi_8 = \begin{pmatrix}0\\b_T\end{pmatrix}.$$

The local basis functions for the pressure component are $\varphi_1, \varphi_2, \varphi_3$. The local matrices then read as

$$A_T = \left[\int_T D\psi_j : D\psi_k \, dx\right]_{j,k=1,\dots,8} \quad \text{and} \quad B_T = \left[-\int_T \varphi_j \operatorname{div} \psi_k \, dx\right]_{\substack{j=1,\dots,3\\k=1,\dots,8}}.$$

We can then compute the entries as follows.

LEMMA 2.14. *The local matrices in the Mini FEM satisfy the following.*

(a) *A_T has the following block structure*

$$A_T = \begin{bmatrix} S & 0 & 0\\ 0 & S & 0\\ 0 & 0 & R\end{bmatrix}$$

for

$$S = \left[\int_T \nabla\varphi_j \cdot \nabla\varphi_k \, dx\right]_{j,k=1,2,3} \quad \textit{and} \quad R = \frac{|T|}{180}\sum_{j=1}^{3} |\nabla\varphi_j|^2 \begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

(b) *B_T has the following block structure*

$$B_T = |T| \left[\begin{array}{c|c} L & \\ L & G \\ L & \end{array}\right]$$

for

$$L = -\frac{1}{3}\begin{bmatrix}\partial_x\varphi_1 & \partial_x\varphi_2 & \partial_x\varphi_3 & \partial_y\varphi_1 & \partial_y\varphi_2 & \partial_y\varphi_3\end{bmatrix} \quad \textit{and} \quad G = \frac{1}{60}\begin{bmatrix}\partial_x\varphi_1 & \partial_y\varphi_1\\ \partial_x\varphi_2 & \partial_y\varphi_2\\ \partial_x\varphi_3 & \partial_y\varphi_3\end{bmatrix}.$$

PROOF. Exercise (recall the integration formula from Problem 1.23). □

Assembling the global system matrix. These local matrices need to be assembled to the global system matrix of the Mini FEM. We need to fix some numbering of degrees of freedom. Let $N_N = \mathrm{card}(\mathcal{N}(\Omega))$ denote the number of interior vertices and $N_T = \mathrm{card}(\mathcal{T})$ denote the number of triangles in the triangulation $\mathcal{T}$ of the 2D domain Ω. With the nodal basis functions $(\lambda_z)_{z\in\mathcal{N}}$ define the following basis functions $\psi_1, \dots, \psi_{2N_N+2N_T}$ for the velocity by

$$(\psi_1, \dots, \psi_{N_N}) = \left[\begin{pmatrix}\varphi_z\\0\end{pmatrix}\right]_{z\in\mathcal{N}(\Omega)}, \quad (\psi_{N_N+1}, \dots, \psi_{2N_N}) = \left[\begin{pmatrix}0\\\varphi_z\end{pmatrix}\right]_{z\in\mathcal{N}(\Omega)},$$
$$(\psi_{2N_N+1}, \dots, \psi_{2N_N+N_T}) = \left[\begin{pmatrix}b_T\\0\end{pmatrix}\right]_{T\in\mathcal{T}},$$
$$(\psi_{2N_N+N_T+1}, \dots, \psi_{2N_N+2N_T}) = \left[\begin{pmatrix}0\\b_T\end{pmatrix}\right]_{T\in\mathcal{T}}.$$

and for the pressure component define

$$(q_1, \dots, q_{N_N}) = [\varphi_z]_{z\in\mathcal{N}(\Omega)}.$$

The global matrices then read as

$$A = \left[\int_\Omega D\psi_j : D\psi_k \, dx\right]_{j,k=1,\dots,2(N_N+N_T)}$$
$$\text{and} \quad B = \left[-\int_\Omega q_j \operatorname{div} \psi_k \, dx\right]_{\substack{j=1,\dots,N_N\\k=1,\dots,2(N_N+N_T)}}.$$

We then observe:

LEMMA 2.15. *The global matrices in the Mini FEM satisfy the following.*

(a) *The global system matrix M of the Mini FEM discretization of the saddle-point formulation has the block structure*

$$M = \begin{bmatrix}A & B^T\\B & 0\end{bmatrix}$$

so that the discrete equation reads as

$$\begin{bmatrix}A & B^T\\B & 0\end{bmatrix}\begin{bmatrix}u_h\\p_h\end{bmatrix} = \begin{bmatrix}F\\0\end{bmatrix}.$$

(b) *The matrix M has a nontrivial kernel, namely globally constant pressure modes.*

PROOF. Exercise. □

So far we did not enforce the discrete pressure variable to belong to $L^2_0(\Omega)$. This constraint is best included via a Lagrange multiplier. Details can be found in the code provided on the course webpage, which follows the implementation outlined above.

The discrete inf-sup condition for the Mini element. We now verify the discrete inf-sup condition for the Mini FEM. The main technical tool is a quasi-interpolation operator.

THEOREM 2.16. *The Mini finite element satisfies the discrete inf-sup condition. On a shape-regular sequence of triangulations $(\mathcal{T}_h)_h$, the discrete inf-sup constant β_h is uniformly bounded from below by some $\beta_0 > 0$.*

PROOF. Given any $q_h \in M_h \subseteq M$, the inf-sup condition for the spaces V and M shows that there exists some $u \in V$ such that $\operatorname{div} u = q_h$ and $\|Du\|_{L^2(\Omega)} \leq \beta^{-1}\|q_h\|_{L^2(\Omega)}$. We set

$$u_h = u_h(q_h) := R_h u + \sum_{T\in\mathcal{T}} \frac{b_T}{\int_T b_T\,dx} \int_T (u - R_h u)\,dx \in V_h.$$

It is shown as an exercise that

$$\|Du_h\|_{L^2(\Omega)} \leq C\|Du\|_{L^2(\Omega)}$$

for some constant $C > 0$. It is furthermore immediate to see that u_h satisfies $\int_T (u_h - u) = 0$ for all $T \in \mathcal{T}$. Therefore, integration by parts reveals

$$b(u_h - u, q_h) = \int_\Omega \operatorname{div}(u_h - u) q_h\,dx = -\int_\Omega (u_h - u)\cdot\nabla q_h\,dx = 0$$

because ∇q_h is piecewise constant. We compute, by plugging in the particular candidate $u_h(q_h)$ in the supremum, that

$$\begin{aligned}\inf_{q_h\in M_h\setminus\{0\}} \sup_{v_h\in V_h\setminus\{0\}} \frac{b(v_h,q_h)}{\|v_h\|_V\|q_h\|_M} &\geq \inf_{q_h\in M_h\setminus\{0\}} \frac{b(u_h(q_h),q_h)}{\|u_h(q_h)\|_V\|q_h\|_M} \\ &= \inf_{q_h\in M_h\setminus\{0\}} \frac{b(u(q_h),q_h)}{\|u_h(q_h)\|_V\|q_h\|_M} \\ &= \frac{1}{C}\inf_{q_h\in M_h\setminus\{0\}} \frac{b(u(q_h),q_h)}{\|u(q_h)\|_V\|q_h\|_M} \geq \frac{\beta}{C}.\end{aligned}$$

The choice $\beta_h := \beta/C$ completes the proof. □

2.2.3. Error estimates. We start by reformulating Theorem 2.7 in the context of saddle-point problems. If we add the two equations of the saddle-point problem (24) we arrive at the equivalent formulation: Find $(u,p) \in V \times M$ such that

$$\mathcal{A}(u,p;v,q) = F(v) + G(q) \quad \text{for all } (v,p) \in V\times M \tag{26}$$

where

$$\mathcal{A}(u,p;v,q) = a(u,p) + b(v,p) + b(u,q)$$

is a continuous bilinear form on $V \times M$. The Brezzi splitting theorem (Theorem 2.10) states that

$$(u,p) \mapsto \mathcal{A}(u,p;\cdot,\cdot)$$

is an isomorphism from $V \times M$ to its dual $V^* \times M^*$ provided a is coercive on the kernel Z of b, and b satisfies the inf-sup condition.
Let $V_h \subseteq V$ and $M_h \subseteq M$ be closed subspaces. The restriction of $\mathcal{A}$ to $V_h \times M_h$ defines a map

$$(u_h,p_h) \mapsto \mathcal{A}(u_h,p_h;\cdot,\cdot)$$

from $V_h \times M_h$ to its dual.

THEOREM 2.17. *Assume $\mathcal{A}$ is an isomorphism. Let $V_h \subseteq V$ and $M_h \subseteq M$ be closed subspaces, let a be coercive on the discrete kernel*

$$Z_h := \{v_h \in V_h : b(v_h,q_h) = 0 \text{ for all } q_h \in M_h\}$$

and let b satisfy the discrete inf-sup condition. Then, given any $F \times G \in V^* \times M^*$, there exists a unique $(u_h,p_h) \in V \times M$ such that

$$\mathcal{A}(u_h,p_h;v_h,q_h) = F(v_h) + G(q_h) \quad \text{for all } (v_h,q_h) \in V\times M.$$

We have the quasi-optimal error estimate

$$\|(u-u_h,p-p_h)\|_{V\times M} \leq (1 + \frac{C_{\mathcal{A}}}{\gamma_h}) \inf_{v_h\in V_h}\inf_{q_h\in M_h} \|(u-v_h,p-q_h)\|_{V\times M}$$

where $C_{\mathcal{A}}$ is the continuity constant of $\mathcal{A}$ and γ_h^{-1} is the continuity constant of the discrete inverse to $\mathcal{A}_h$.

PROOF. From Theorem 2.10 applied to the discrete spaces, we deduce that $\mathcal{A}_h$ is an isomorphism from $V_h \times M_h$ to its dual. With the continuity constant $\gamma_h > 0$ of the inverse operator we have in particular that

$$\gamma_h \|(v_h, q_h)\|_{V_h \times M_h} \leq \|\mathcal{A}_h(v_h, q_h)\|_{V_h^* \times M_h^*}.$$

The existence and uniqueness of the approximate solution (u_h, p_h) as well as the error estimate follow from the abstract bound of Theorem 2.7. □

WARNING 2.18. In general we expect $Z_h \not\subseteq Z$, i.e., the kernel spaces are not nested.

Checking the discrete inf-sup condition can be a difficult task. In the proof of Theorem 2.16 we have constructed a bounded operator from V to V_h, $u \mapsto u_h$, with the property $b(u_h - u, q_h) = 0$ for all $q_h \in M_h$. Such an operator is called *Fortin operator*. Constructing a Fortin operator often is a suitable method for verifying the inf-sup condition.

LEMMA 2.19 (Fortin criterion). *Let the bilinear form $b : V \times M \to \mathbb{R}$ satisfy the inf-sup condition. Assume that for closed subspaces $V_h \subseteq V$ and $M_h \subseteq M$ there exists a bounded linear map $\Pi_h : V \to V_h$ with the property that*

$$b(v - \Pi_h v, q_h) = 0 \quad \textit{for all } q_h \in M_h.$$

Then, the discrete inf-sup constant is satisfied with a constant proportional to the inverse of the continuity constant of Π_h.

PROOF. We repeat the argument from Theorem 2.16 in this abstract framework. The inf-sup condition for V and M and the properties of Π_h show for any $q_h \in M_h \subseteq M$ that

$$\begin{aligned}\beta\|q_h\|_M \leq \sup_{v \in V\setminus\{0\}} \frac{b(v, q_h)}{\|v\|_V} &= \sup_{v \in V\setminus\{0\}} \frac{b(\Pi_h v, q_h)}{\|v\|_V} \\ &\leq C \sup_{\substack{v \in V\setminus\{0\} \\ \Pi_h v \neq 0}} \frac{b(\Pi_h v, q_h)}{\|\Pi_h v\|_V} \leq C \sup_{v_h \in V_h\setminus\{0\}} \frac{b(v_h, q_h)}{\|v_h\|_V}.\end{aligned}$$

Here, we denoted the continuity constant of Π_h by C. □

Let us conclude the study of the Stokes equations by summarizing our findings for the Mini finite element as an error estimate.

THEOREM 2.20. *The mini finite element discretization (u_h, p_h) to the Stokes equations satisfies*

$$\begin{aligned}&\|D(u - u_h)\|_{L^2(\Omega)} + \|p - p_h\|_{L^2(\Omega)} \\ &\quad \leq C\Big(\inf_{v_h \in [S_0^1(\mathcal{T}) + \mathcal{B}(\mathcal{T})]^2} \|D(u - v_h)\|_{L^2(\Omega)} + \inf_{q_h \in P_0(\mathcal{T}) \cap L_0^2(\Omega)} \|p - q_h\|_{L^2(\Omega)}\Big)\end{aligned}$$

for some constant C that is independent of the mesh size.

REMARK 2.21. In case that Ω is convex, it is known that additionally we have $u \in [H^2(\Omega)]^2$ and $p \in H^1(\Omega)$ with

$$\|D^2 u\|_{L^2(\Omega)} + \|\nabla p\|_{L^2(\Omega)} \leq C\|f\|_{L^2(\Omega)}.$$

Together with suitable (quasi-)interpolation estimates we therefore conclude that the error of the Mini FEM is bounded by some $\tilde{C}h\|f\|_{L^2(\Omega)}$ on convex domains and thus converges at order h. ♦

2.3. Variational problems in H(div)

2.3.1. Duality in Hilbert spaces. For a Hilbert space Y, the Riesz representation theorem establishes an isometry between Y and its dual. That is, we can identify any $y \in Y$ with the linear functional $\langle y, \cdot \rangle_Y$. For example, any element of $L^2(\Omega)^*$ can be represented by $\int_\Omega g \cdot dx$ for some $f \in L^2(\Omega)$. Such identifications are very common and culminate in statements like "Y is its own dual", but some care is necessary when working with them. In particular, when working with more than one Hilbert space, it must be clear with respect to which space we take this identification.

Definition 2.22 (Gelfand triplet). Let X, Y be Hilbert spaces where X is densely embedded in Y. We know (Exercise 2.5) that Y^* is then densely embedded in X^*. After identifying Y with Y^* we therefore have the chain of embeddings

$$X \to Y \to X^*.$$

This is called a *Gelfand triplet* and Y is called the *pivot space*. ♦

We proceed with the most prominent example in our lecture, which is related to Sobolev spaces.

Example 2.23 (embedding in negative Sobolev spaces). Given a bounded polyhedral Lipschitz domain Ω, recall the spaces $H^1(\Omega)$ $H^1_0(\Omega)$. As usual, we write $H^1_0(\Omega)^* = H^{-1}(\Omega)$. We know that the embedding $H^1(\Omega) \subseteq L^2(\Omega)$ is dense and, after identifying $L^2(\Omega)$ with its dual, we arrive at the inclusions

$$H^1(\Omega) \subseteq L^2(\Omega) \subseteq H^1(\Omega)^* \quad \text{and} \quad H^1_0(\Omega) \subseteq L^2(\Omega) \subseteq H^{-1}(\Omega).$$

Warning 2.24 (pivot space). In stating such inclusions, it is of paramount importance to specify the pivot space. Anything else will be prone to heavy mistakes.

Example 2.25 (Dirichlet Laplacian). We know the well-posedness of the weak Poisson equation $-\Delta u = f$ in Ω subject to the boundary condition $u|_{\partial\Omega} = 0$. The solution $u \in H^1_0(\Omega)$ is the Riesz representative of the functional $f \in H^{-1}(\Omega)$. If there exists $T_f \in L^2(\Omega)$ with $f = \int_\Omega T_f \cdot dx$, we use the identification of $L^2(\Omega)$ with itself to interpret the inclusion $f \in H^{-1}(\Omega)$ and say that "f is an L^2 function". Without specifying the underlying identification, the statement obviously makes no sense because the elements $H^{-1}(\Omega)$ are not functions. Note that not every element of $H^{-1}(\Omega)$ may have an L^2 representation.

Example 2.26 (Neumann Laplacian). The Neumann Laplacian problem is to find $u \in H^1(\Omega)$ with $\int_\Omega u\,dx = 0$ and

$$-\Delta u = f \quad \text{in } \Omega \quad \text{and} \quad \partial u/\partial \nu = 0 \quad \text{on } \partial\Omega$$

for the outer unit normal ν. A necessary compatibility condition of f comes from the divergence theorem and reads as

$$\int_\Omega f\,dx = -\int_\Omega \Delta u\,dx = -\int_{\partial\Omega} \partial u/\partial\nu\,ds = 0.$$

We denote $Z := H^1(\Omega)/\mathbb{R} := \{v \in H^1(\Omega) : \int_\Omega v\,dx = 0\}$ and recall the weak formulation of the problem, namely: find $u \in Z$ such that

$$\int_\Omega \nabla u \cdot \nabla v\,dx = \int_\Omega f v\,dx.$$

This weak problem is, of course, well posed for any $f \in L^2(\Omega)$. The constraint on the average of f is not needed. The point is that any constant function f will result in $\int_\Omega f v\,dx = 0$ for any $v \in Z$ and therefore represents the zero element of Z^*. Indeed, the inclusion $Z \subseteq L^2(\Omega)$ is not dense. The correct pivot space in the

Gelfand triple is therefore the space $L^2_0(\Omega)$ of L^2 functions with vanishing average. This resembles the above compatibility condition.

EXAMPLE 2.27 (Neumann Laplacian as a saddle-point problem). The Neumann Laplacian problem can be posed as a variational problem over $V := H^1(\Omega)$. We denote by $M \eqsim \mathbb{R}$ the space of constant functions and introduce the operator $B : V \to M$, $v \mapsto \int_\Omega v \cdot dx$. Denoting by A the gradient inner product, we see that A is coercive on Z (Poincaré's inequality) and that B trivially satisfies the inf-sup condition. Therefore, there exists a constant $p \in M$ such that

$$\begin{aligned} \int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega pv, \, dx &= \int_\Omega fv \, dx && \text{for all } v \in V \\ \int_\Omega uq \, dx &= 0 && \text{for all } q \in M. \end{aligned}$$

It is easy to see that $p = \fint_\Omega f \, dx$ equals the average of f, which conforms to the fact that only the projection of f to Z has an effect on u.

Given a bounded polyhedral Lipschitz domain Ω, we already know the spaces $H^1(\Omega)$ and $H^1_0(\Omega)$. The trace theorem teaches us that a function $v \in H^1(\Omega)$ admits boundary values $v|_{\partial\Omega} \in L^2(\partial\Omega)$ in the sense of traces. That is, there exists a linear and continuous operator $T : H^1(\Omega) \to L^2(\partial\Omega)$ that coincides with the usual restriction to the boundary when applied to continuously differentiable functions. We recall that $H^1_0(\Omega)$ is the closure of $C^\infty_c(\Omega)$ unter the $H^1(\Omega)$ norm and can be characterized as the subspace of $H^1(\Omega)$ of functions with vanishing trace. The range of the trace operator is customarily denoted by

$$H^{1/2}(\partial\Omega) := T(H^1(\Omega)).$$

(The reason for this notation will become clear later in this lecture.) It is equipped with the minimal extension norm

$$\|g\|_{H^{1/2}(\partial\Omega)} := \inf_{v \in H^1(\Omega) : Tv = g} \|v\|_{H^1(\Omega)}.$$

The minimal extension is the solution to an elliptic boundary value problem (see Exercise 2.18 for a similar computation). We denote by $H^{-1/2}(\partial\Omega)$ the dual space of $H^{1/2}(\partial\Omega)$. The norm in that space is, as usual, defined as

$$\|q\|_{H^{-1/2}(\partial\Omega)} = \sup_{v \in H^{1/2}(\partial\Omega)} \frac{\langle q, v \rangle}{\|v\|_{H^{1/2}(\partial\Omega)}}.$$

We have the Gelfand triplet

$$H^{1/2}(\partial\Omega) \subseteq L^2(\partial\Omega) \subseteq H^{-1/2}(\partial\Omega),$$

for which we will later verify that the embedding is indeed dense.
If we now define by $H^{1/2}(\Gamma)$ for some $\Gamma \subseteq \partial\Omega$ the range of the trace operator restricted to Γ we are not working on a closed manifold anymore. Formal integration by parts with a function $v \in H^{1/2}(\Gamma)$ will cause boundary terms unless it admits an extension by zero to a function $\tilde{v}$ in $H^{1/2}(\partial\Omega)$. The space of such functions is defined as

$$\tilde{H}^{1/2}(\Gamma) := \{v \in H^{1/2}(\Gamma) : \tilde{v} \in H^{1/2}(\partial\Omega)\}.$$

We observe that in general $H^{-1/2}(\Gamma)$ and $(\tilde{H}^{1/2}(\Gamma))^*$ are different spaces. This is a delicate issue that we will discuss in more detail.

2.3.2. The space H(div). If we consider the Dirichlet problem $-\Delta u = f$ for the Laplacian with $f \in L^2(\Omega)$, we notice that $\sigma := \nabla u$ is an element of $[L^2(\Omega)]^n$. But we know more, namely

$$\int_\Omega \sigma \cdot \nabla v \, dx = -\int_\Omega f v \, dx \quad \text{for all } v \in C_c^\infty(\Omega).$$

That is, σ is in L^2 and possesses a *weak divergence* in L^2. The space of such vector fields is denoted by

$$H(\operatorname{div}, \Omega) := \left\{ \sigma \in [L^2(\Omega)]^n : \exists f \in L^2(\Omega) \; \forall v \in C_c^\infty(\Omega) \int_\Omega \sigma \cdot \nabla v \, dx = -\int_\Omega f v \, dx \right\}.$$

The weak divergence is then denoted by $\operatorname{div} \sigma = f$. The space is endowed with the norm

$$\|v\|_{H(\operatorname{div},\Omega)} := \sqrt{\|v\|^2_{L^2(\Omega)} + \|\operatorname{div} v\|^2_{L^2(\Omega)}}.$$

One can show that $H(\operatorname{div}, \Omega)$ is the closure of the smooth vector fields (up to the boundary) with respect to the norm $\|\cdot\|_{H(\operatorname{div},\Omega)}$.
Of course, any vector field whose components all belong to $H^1(\Omega)$ automatically belong to $H(\operatorname{div}, \Omega)$. But $H(\operatorname{div})$ fields are more general. For example (see Exercise 2.21), a piecewise polynomial vector field with respect to a triangulation need not be globally continuous to belong to that space. It suffices that it does not jump across any face in the direction normal to that face.
Functions from $H(\operatorname{div})$ have traces in a certain sense. Integration by parts shows (for sufficiently smooth functions) that

$$\int_{\partial\Omega} \varphi \tau \cdot \nu \, dx = \int_\Omega \varphi \operatorname{div} \tau \, dx + \int_\Omega \tau \cdot \nabla \varphi \, dx \le \|\tau\|_{H(\operatorname{div},\Omega)} \|\varphi\|_{H^1(\Omega)}.$$

This means that the normal trace, assigning $\tau \cdot \nu|_{\partial\Omega}$ to any τ, is a bounded linear functional on $H^{1/2}(\partial\Omega)$.

Example 2.28 (inhomogeneous Neumann problem). Given $g \in H^{-1/2}(\partial\Omega)$, the weak form of the Neumann problem $-\Delta u + u = f$, $\partial u/\partial\nu = g$ seeks $u \in H^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v \, dx + \int_\Omega uv \, dx = \int_\Omega f v \, dx + \int_{\partial\Omega} g v \, ds \quad \text{for all } v \in H^1(\Omega).$$

It is well posed and its unique solution u satisfies $\nabla u \cdot \nu = g$ on $\partial\Omega$ as an identity of elements in $H^{-1/2}(\partial\Omega)$. We therefore see that the normal trace is surjective onto that space.

Previously, we could easily restrict $H^{1/2}$ functions from $\partial\Omega$ to a subset $\Gamma \subseteq \partial\Omega$. This is not possible for elements from $H^{-1/2}(\partial\Omega)$. Indeed, by our above interpretation of the normal trace through integration by parts, we think of an identity

$$\int_\Gamma \varphi \tau \cdot \nu \, dx = \int_{\partial\Omega} \hat{\varphi} \tau \cdot \nu \, dx$$

where $\hat{\varphi}$ is the zero extension of φ. But that extension need not belong to $H^{1/2}(\partial\Omega)$. All subsequent computations from above then will make no sense any more.

Example 2.29. For $g \in H^{-1/2}(\partial\Omega)$ and a (generic) subset $\Gamma \subseteq \partial\Omega$, the integral $\int_\Gamma g \, ds$ is not well defined because the constant 1 over Γ is not in $H^{1/2}(\partial\Omega)$ when continued by 0. We will study this in more detail later, but for the moment we consider another example.

Example 2.30 (taken from §2.5.1 of [**BBF13**]). We know that in two dimensions the function $u(x) = \log(|\log(|x|)|)$ belongs to $H^1(\Omega)$ and so its trace belongs to $H^{1/2}(\partial\Omega)$. For simpler computations, we take Ω to be the quarter segment $\Omega =$

$\{x_1 > 0, x_2 > 0, |x| \leq 1/\exp(1)\}$. The tangential derivative of u along $\partial\Omega$ then belongs to $H^{-1/2}(\partial\Omega)$ (this will be proven later in the lecture) and is denoted by g. By direct computation, we see that $\int_{\partial\Omega} g\,ds$ is finite, but $\int_\Gamma g\,ds$ for $\Gamma = \{x_2 = 0\} \cap \partial\Omega$ is not.

2.3.3. Mixed finite elements for Poisson's equation. In Poisson's equation we introduce an additional vector variable σ and set

$$\sigma = \nabla u, \qquad -\operatorname{div}\sigma = f.$$

For the variable σ above we require $\sigma \in H(\operatorname{div},\Omega)$ and

$$\int_\Omega \operatorname{div}\sigma\, v\,dx = -\int_\Omega fv\,dx \quad \text{for all } v \in L^2(\Omega).$$

The relation $\sigma = \nabla u$ and integration by parts reveal for any $\tau \in H(\operatorname{div},\Omega)$ that

$$\int_\Omega \sigma\cdot\tau\,dx = -\int_\Omega \operatorname{div}\tau u\,dx + \int_{\partial\Omega} u\tau\cdot\nu\,ds.$$

Assuming a homogeneous Dirichlet boundary condition for u we $\sigma \in H(\operatorname{div},\Omega)$ and $u \in L^2(\Omega)$ such that

$$\begin{aligned}\int_\Omega \sigma\cdot\tau\,dx + \int_\Omega \operatorname{div}\tau u\,dx = 0 \qquad & \text{for all } \tau \in H(\operatorname{div},\Omega),\\ \int_\Omega \operatorname{div}\sigma\, v\,dx = -\int_\Omega fv\,dx \qquad & \text{for all } v \in L^2(\Omega).\end{aligned}$$

In this way we have formulated Poisson's equation as a saddle-point problem. This formulation is referred to as *mixed formulation*. As an exercise it is shown that the system satisfies the properties from Brezzi's splitting theorem and is therefore well-posed. We remark that we have explicitly imposed the $H(\operatorname{div},\Omega)$ regularity for the vector variable but now merely ask u to belong to $L^2(\Omega)$. The property that σ is the weak gradient of u is implicitly contained in the first row of the system.
We want to identify appropriate finite element spaces leading to a stable discretization of the mixed Laplacian. Since $L^2(\Omega)$ functions do not require any continuity, a reasonable choice is to discretize it by the subspace $P_0(\mathcal{T})$ of piecewise constant (possibly discontinuous) functions with respect to a regular simplicial triangulation $\mathcal{T}$. For piecewise polynomial discretizations of $H(\operatorname{div},\Omega)$ we have seen in Problem 2.21 that for each face of the triangulation the component of the piecewise polynomial vector field must be continuous in the normal direction with respect to the face. We thus will use the normal directions at the faces as the degrees of freedom. For simplicity, we restrict ourselves to two space dimensions but remark that an analogous reasoning works in any dimension. We begin with the construction on a single triangle. We set

$$RT_0(T) := \{v \in [L^2(T)]^2 : v(x) = \begin{pmatrix} a \\ b \end{pmatrix} + cx \text{ for } a,b,c \in \mathbb{R}\}.$$

The vector fields of $RT_0(T)$ belong to a subset of the vector fields that are affine in each component. Obviously $\dim RT_0(T) = 3$. For the standard P_1 finite element, the degrees of freedom were the point evaluations at the vertices and we worked with the nodal basis of hat functions. Since here we want to enforce continuity of the normal component across edges we seek a basis $(\psi_E)_{E\in\mathcal{E}(T)}$, where $\mathcal{E}(T)$ is the set of edges of T, such that

$$\fint_F \psi_E\cdot\nu_F\,dx = \begin{cases} 1 & \text{if } E = F \\ 0 & \text{else.}\end{cases} \tag{27}$$

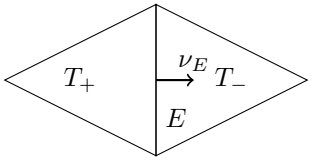


FIGURE 1. Convention for the edge normal.

Here, ν_F is the outer normal vector of T restricted to the edge F. This property is achieved by the following definition

$$\psi_{T,E}(x) := \frac{|E|}{2|T|}(x - P_E)$$

where P_E is the vertex of T opposite to E. The proof of (27) is left as an exercise.

REMARK 2.31 (finite element in the sense of Ciarlet). In the foregoing discussion, we have seen that we can uniquely determine functions from a finite-dimensional space of functions over T by linear functionals that need not be point evaluations (as it would be the case for the usual Lagrange basis of polynomials). Following the reasoning of Ciarlet [**Cia78**], one can abstractly define a *finite element* as a triplet $(T, \mathcal{P}, \mathcal{L})$ consisting of a bounded Lipschitz domain T of $\mathbb{R}^n$ (the element domain), a finite-dimensional space $\mathcal{P}$ of functions over T (the *shape functions*), and a set $\mathcal{L}$ of linear functionals over $\mathcal{P}$ that forms a basis of P^* (the *node functionals*). It is an exercise to verify that $(T, RT_0(T), \{⨍_E \bullet \cdot \nu_T \, ds : E \in \mathcal{E}(T)\})$ is a finite element. ♦

Globally, we then define

$$RT_0(\mathcal{T}) := \{v \in H(\mathrm{div}, \Omega) : \forall T \in \mathcal{T} \; v|_T \in RT_0(T)\}.$$

This space is called the Raviart–Thomas finite element space. We have seen that it consists of all vector fields that are in $RT_0(T)$ for every triangle T and that are normal-continuous across each edge. Given any interior edge E, we fix a normal vector. For the two neighbouring triangles T_+ and T_- this vector then points inwards to one of them and outwards to the other one. We use the convention that

$$\nu_E = \nu_{T_+} \quad \text{and} \quad \nu_E = -\nu_{T_-}$$

that is, ν_E is the outward pointing normal to T_+. This is graphically illustrated in Figure 1. If E is a boundary edge, we define $T_- = \emptyset$.
The functions

$$\psi_E(x) = \begin{cases} \psi_{T_+,E}(x) & \text{if } x \in T_+ \\ -\psi_{T_-,E}(x) & \text{if } x \in T_- \\ 0 & \text{else} \end{cases}$$

then form a global basis of $RT_0(\mathcal{T})$.

LEMMA 2.32. *The functions $(\psi_E)_{E\in\mathcal{E}}$ form a basis of $RT_0(\mathcal{T})$. They satisfy $⨍_F \psi_E \cdot \nu_F \, dx = \delta_{EF}$.*

PROOF. Exercise. □

The Raviart–Thomas space has a canonical interpolation operator, which reads for any sufficiently smooth vector field τ

$$I_{RT}\tau = \sum_{E\in\mathcal{E}} ⨍_E \tau \cdot \nu_E \, ds \psi_E.$$

By construction, it satisfies the conservation property

$$\int_E I_{RT}\tau \cdot \nu_E \, ds = \int_E \tau \cdot \nu_E \, ds \quad \text{for any } E \in \mathcal{E}.$$

We will see that this operator is not well defined for functions in $H(\mathrm{div}, \Omega)$ but requires further regularity of τ. A sufficient criterion for $I_{RT}\tau$ to exist is for instance $\tau \in [H^1(\Omega)]^2$ because traces along edges are well defined due to the trace theorem. The following result shows H^1 stability of I_{RT}.

THEOREM 2.33. *The Raviart–Thomas interpolation is stable with respect to the H^1 norm in the following sense. There exists a constant $C_{I_{RT}}$ only dependent on the shape regularity of $\mathcal{T}$ such that*

$$\|I_{RT}v\|_{H^1(\Omega)} \leq C_{I_{RT}} \|v\|_{H^1(\Omega)} \quad \textit{for all } v \in [H^1(\Omega)]^2.$$

PROOF. The restriction of $I_{RT}v$ to a triangle K can be written in terms of the basis expansion as follows

$$I_{RT}v|_K = \sum_{E \in \mathcal{E}(K)} \fint_E v \cdot \nu_E \, ds \psi_E.$$

A direct computation with the shape regularity shows for the basis function that $\|\psi_E\|_{L^2(K)} \lesssim h_K$. Similarly, $\|D\psi_E\|_{L^2(K)} \lesssim 1$. We recall the trace inequality and compute for the coefficient in front of ψ_E that

$$|\fint_E v \cdot \nu_E| \leq |E|^{-1/2} \|v\|_{L^2(E)} \lesssim h_K^{-1} \|v\|_{L^2(K)} + \|Dv\|_{L^2(K)}.$$

We use the triangle inequality and compute

$$\|I_{RT}v\|_{L^2(K)} \leq \sum_{E \in \mathcal{E}(K)} |\fint_E v \cdot \nu_E \, ds| \|\psi_E\|_{L^2(\Omega)} \lesssim \|v\|_{H^1(K)}.$$

In order to bound the gradient, we observe that $DI_{RT}v = DI_{RT}(v - \fint_K v \, dx)$ for the constant $\fint_K v \, dx$ (component-wise integral mean) because I_{RT} conserves constants (exercise). We then compute with trace and Poincaré inequalities that

$$\begin{aligned}
\|DI_{RT}v\|_{L^2(K)} &= \|DI_{RT}(v - \fint_K v \, dx)\|_{L^2(\Omega)} \\
&\leq \sum_{E \in \mathcal{E}(K)} |\fint_E (v - \fint_K v \, dx) \cdot \nu_E \, ds| \|D\psi_E\|_{L^2(\Omega)} \\
&\lesssim h_K^{-1} \|(v - \fint_K v \, dx)\|_{L^2(K)} + \|\nabla v\|_{L^2(K)} \lesssim \|v\|_{H^1(K)}.
\end{aligned}$$

Note that the constant in the Poincaré inequality scales like h_K. The claimed bound on the $H^1(\Omega)$ norm follows from using this local argument on each element domain. □

The following so-called *commuting diagram property* is of particular importance. We denote by $\Pi_0 : L^2(\Omega) \to P_0(\mathcal{T})$ the L^2 projection on piecewise constants. It has the following representation (exercise)

$$(\Pi_0 q)|_T = \fint_T q \, dx \quad \text{for all } q \in L^2(\Omega) \text{ and all } T \in \mathcal{T}.$$

For vector variables, we use the same symbol Π_0 to denote the component-wise L^2 projection on $[P_0(\mathcal{T})]^2$.

LEMMA 2.34 (commuting diagram property). *The Raviart–Thomas interpolation* $I_{RT} : [H^1(\Omega)]^2 \to \mathbb{R}T_0(\mathcal{T})$ *satisfies*

$$\operatorname{div} I_{RT} v = \Pi_0 \operatorname{div} v.$$

In other words, the diagram

$$\begin{array}{ccc} [H^1(\Omega)]^2 & \xrightarrow{\operatorname{div}} & L^2(\Omega) \\ \Big\downarrow{\scriptstyle I_{RT}} & & \Big\downarrow{\scriptstyle \Pi_0} \\ RT_0(\mathcal{T}) & \xrightarrow{\operatorname{div}} & P_0(\mathcal{T}) \end{array}$$

commutes.

PROOF. Let $v \in [H^1(\Omega)]^2$. The divergence theorem shows for any $T \in \mathcal{T}$ with outer unit normal ν that

$$\int_T \operatorname{div} I_{RT} v \, dx = \int_{\partial T} I_{RT} v \cdot \nu_T \, ds = \sum_{E \in \mathcal{E}(T)} \int_E I_{RT} v \cdot \nu|_E \, ds.$$

For any edge $E \in \mathcal{E}(T)$, the operator I_{RT} conserves the integral of $v \cdot \nu|_E$. Thus

$$\sum_{E \in \mathcal{E}(T)} \int_E I_{RT} v \cdot \nu|_E \, ds = \sum_{E \in \mathcal{E}(T)} \int_E v \cdot \nu|_E \, ds = \int_{\partial T} v \cdot \nu \, ds = \int_T \operatorname{div} v \, ds$$

where we used again the divergence theorem. We combine the above two chains of identities and divide by the area of T to obtain

$$⨍_T \operatorname{div} I_{RT} v \, dx = ⨍_T \operatorname{div} v \, ds.$$

The left integrals simply equals $\operatorname{div} I_{RT} v$ because the integrand is constant on T. The assertion follows with the above representation of Π_0 as the piecewise integral mean. □

Let us now turn to the discretization of the mixed Laplacian. The mixed finite element approximation seeks $(\sigma_h, u_h) \in RT_0(\mathcal{T}) \times P_0(\mathcal{T})$ such that

$$\begin{aligned} \int_\Omega \sigma_h \cdot \tau_h \, dx - \int_\Omega \operatorname{div} \tau_h u_h \, dx = 0 \qquad & \text{for all } \tau_h \in RT_0(\mathcal{T}), \\ \int_\Omega \operatorname{div} \sigma_h \, v_h \, dx = -\int_\Omega f v_h \, dx \qquad & \text{for all } v_h \in P_0(\mathcal{T}). \end{aligned}$$

THEOREM 2.35. *Given any* $f \in L^2(\Omega)$*, there is a unique solution* $(\sigma_h, u_h) \in RT_0(\mathcal{T}) \times P_0(\mathcal{T})$ *to the discrete mixed system. We have the error estimate*

$$\begin{aligned} &\|\sigma - \sigma_h\|_{H(\operatorname{div},\Omega)} + \|u - u_h\|_{L^2(\Omega)} \\ &\quad \le C\big(\inf_{\tau_h \in RT_0(\mathcal{T})} \|\sigma - \tau_h\|_{H(\operatorname{div},\Omega)} + \inf_{v_h \in P_0(\mathcal{T})} \|u - v_h\|_{L^2(\Omega)} \big) \end{aligned}$$

for some constant C*.*

PROOF. It suffices to prove the requirements from Brezzi's splitting theorem. The error estimate then follows from the abstract error estimate for the Galerkin method. For the proof of coercivity of the form

$$a(\sigma_h, \tau_h)$$

on the kernel Z_h, we first note that any $\tau_h \in Z_h$ satisfies by definition

$$\int_\Omega \operatorname{div} \tau_h v_h \, dx = 0 \quad \text{for all } v_h \in P_0(\mathcal{T}).$$

But since $\operatorname{div}\tau_h \in P_0(\mathcal{T})$, we see that $\operatorname{div}\tau_h = 0$ pointwise in Ω. Therefore

$$a(\tau_h,\tau_h) = \|\tau_h\|^2_{L^2(\Omega)} = \|\tau_h\|^2_{L^2(\Omega)} + \|\operatorname{div}\tau_h\|^2_{L^2(\Omega)} = \|\tau_h\|^2_{H(\operatorname{div},\Omega)},$$

which implies coercivity of a in $RT_0(\mathcal{T}) \subseteq H(\operatorname{div},\Omega)$.
Let us prove the inf-sup condition for the form

$$b(\tau_h, v_h) := \int_\Omega \operatorname{div}\tau_h\, v_h\, dx.$$

Let any $v_h \in P_0(\mathcal{T})$ be given. In case that Ω is not convex, we increase the domain to a larger convex domain $\hat\Omega$ by adding suitable triangles, and we extend v_h by zero to a function $\hat f \in L^2(\hat\Omega)$. On $\hat\Omega$ we then solve the weak form of the Dirichlet problem $\Delta\hat w = \hat f$ for some $\hat w \in H^1_0(\hat\Omega)$. From the H^2 regularity on convex domains (Part I of this lecture) we deduce that $\hat w \in H^2(\hat\Omega)$ with

$$\|\hat w\|_{H^2(\Omega)} \le C_{\mathrm{reg}}\|v_h\|_{L^2(\Omega)}, \quad \nabla\hat w|_\Omega \in [H^1(\Omega)]^2, \quad \text{and} \quad \operatorname{div}\nabla\hat w = v_h \text{ in } \Omega.$$

Since $\nabla\hat w$ in H^1, its Raviart–Thomas interpolation is well defined and satisfies, due to the commuting diagram property,

$$\operatorname{div} I_{RT}\nabla\hat w = \Pi_0 \operatorname{div}\nabla\hat w = \Pi_0(v_h) = v_h.$$

We furthermore have a bound on the $H(\operatorname{div},\Omega)$ norm

$$\begin{aligned}\|I_{RT}\nabla\hat w\|^2_{H(\operatorname{div},\Omega)} &= \|I_{RT}\nabla\hat w\|^2_{L^2(\Omega)} + \|v_h\|_{L^2(\Omega)} \\ &\lesssim \|\nabla\hat w\|^2_{H^1(\Omega)} + \|v_h\|^2_{L^2(\Omega)} \lesssim \|v_h\|^2_{L^2(\Omega)}.\end{aligned}$$

We then compute

$$\begin{aligned}\sup_{\tau_h\in RT_0(\mathcal{T})\setminus\{0\}} \frac{b(\tau,v_h)}{\|\tau\|_{H(\operatorname{div})}\|v_h\|_{L^2(\Omega)}} &\ge \frac{b(I_{RT}\nabla\hat w, v_h)}{\|I_{RT}\nabla\hat w\|_{H(\operatorname{div})}\|v_h\|_{L^2(\Omega)}} \\ &= \frac{\|v_h\|^2_{L^2(\Omega)}}{\|I_{RT}\nabla\hat w\|_{H(\operatorname{div})}\|v_h\|_{L^2(\Omega)}} \gtrsim 1.\end{aligned}$$

This proves the inf-sup condition with a constant that only depends on the shape regularity. □

Corollary 2.36. *If the solution to the Poisson equation satisfies $u \in H^1_0(\Omega)\cap H^2(\Omega)$, then*

$$\|\sigma-\sigma_h\|_{H(\operatorname{div},\Omega)} + \|u-u_h\|_{L^2(\Omega)} \le h\|D^2u\|_{L^2(\Omega)} + \|f-\Pi_0 f\|_{L^2(\Omega)}.$$

Proof. This follows from the interpolation error estimate and the piecewise Poincaré inequality. □

2.3.4. Selected aspects.

Example 2.37 (mixed BVP). Given a disjoint partition $\partial\Omega = \Gamma_D \cup \Gamma_N$ into a Dirichlet and a Neumann boundary, we consider the mixed boundary value problem $-\Delta u = f$ subject to $u|_{\Gamma_D} = u_D$ and $u|_{\Gamma_N} = 0$ for some given $u_D \in H^{1/2}(\Gamma_D)$. For simplicity we assume Γ_D to have positive surface measure. We introduce the space

$$H_N(\operatorname{div},\Omega) := \{\tau \in H(\operatorname{div},\Omega) : \tau\cdot\nu|_{\Gamma_N} = 0\}$$

and obtain the mixed formulation of the boundary value problem: Find $\sigma \in H_N(\operatorname{div},\Omega)$ and $u \in L^2(\Omega)$ such that

$$\begin{aligned}\int_\Omega \sigma\cdot\tau\, dx + \int_\Omega \operatorname{div}\tau\, u\, dx &= \langle \tau\cdot\nu, u_D\rangle \qquad \text{for all } \tau \in H_N(\operatorname{div},\Omega), \\ \int_\Omega \operatorname{div}\sigma\, v\, dx &= -\int_\Omega f v\, dx \qquad \text{for all } v \in L^2(\Omega).\end{aligned}$$

Note that the Neumann condition enters as an essential boundary condition, while the Dirichlet condition is imposed weakly and appears on the right-hand side. This situation is "dual" to the usual formulation of the boundary value problem studied earlier. Inhomogeneous Neumann conditions have to be imposed in an essential way.

Transformation properties. When working with the Sobolev space $H^1(\Omega)$, for the usual Lagrange elements we know the affine equivalence to a reference element. We can parametrize T via an affine diffeomorphism $\Phi : \hat{T} \to T$ and know that the nodal functionals (point evaluations) are conserved under this transform. We also know the important relation $\nabla v = (D\Phi)^{-\top} \nabla \hat{v} \circ \Phi^{-1}$. In $H(\mathrm{div})$ problems, the situation is different because Φ does not map normal vectors to normal vectors and, thus, does not conserve the degrees of freedom. It turns out (and is well known from the theory of differential forms) that the right transform is the *pullback*, also known as *contravariant transform* or *Piola transform*. For an element $\hat{x} \in \hat{T}$ it acts on a vector field $\hat{q}$ as follows

$$x := \Phi(\hat{x}) \quad \text{and} \quad q(x) := |\det D\Phi(\hat{x})|^{-1} D\Phi(\hat{x}) \hat{q}(\hat{x}).$$

For affine Φ, the object $D\Phi$ can be thought of as a constant matrix, henceforth denoted by B. It is possible to verify that the normal vector ν to ∂T and the normal $\hat{\nu}$ to $\partial \hat{T}$ transform as

$$\nu(x) = \frac{1}{|B^{-\top} \hat{\nu}(\hat{x})|} B^{-\top} \hat{\nu}(\hat{x}),$$

see Exercise 2.31. We furthermore have:

LEMMA 2.38. *Let $q \in H(\mathrm{div}, T)$ be the Piola transform of $\hat{q}$ and $v \in H^1(T)$ be the affine transform of $\hat{v}$. Then*

$$\int_T \mathrm{div}\, q v \, dx = \int_{\hat{T}} \mathrm{div}\, \hat{q} \hat{v} \, dx, \quad \int_T q \cdot \nabla v \, dx = \int_{\hat{T}} \hat{q} \cdot \nabla \hat{v} \, dx,$$
$$\int_{\partial T} q \cdot \nu v \, ds = \int_{\partial \hat{T}} \hat{q} \cdot \hat{\nu} \hat{v} \, ds.$$

PROOF. Exercise 2.32. □

2.3.5. Error estimate in the H^{-1} norm. For the standard FEM, the Aubin–Nitsche trick can be used to establish an improved convergence rate for the L^2 norm of the error compared to the energy norm (the H^1 seminorm). For the mixed FEM, this is obviously impossible because it approximates u in L^2 with piecewise constants. This approximation will be of order h in case of full regularity, but not better. We first study the projected error $\Pi_h(u_h - u)$ and see that it exhibits a superconvergence phenomenon. For simplicity we state it on a convex domain where we know that the Poisson problem is H^2 regular.

LEMMA 2.39. *Let $\Omega \subseteq \mathbb{R}^n$ be an open, bounded, convex polytope. Let $(\sigma, u) \in H(\mathrm{div}, \Omega) \times L^2(\Omega)$ solve the mixed system for the Poisson problem with right-hand side in $L^2(\Omega)$ and homogeneous Dirichlet boundary conditions. Let (σ_h, u_h) denote approximation from the discrete mixed system. Then the projected error satisfies*

$$\|\Pi_h(u_h - u)\|_{L^2(\Omega)} \lesssim h \|\sigma - \sigma_h\|_{H(\mathrm{div},\Omega)}.$$

PROOF. Let $(\eta, w) \in H(\mathrm{div}, \Omega) \times L^2(\Omega)$ denote the solution to the mixed system with right-hand side $\Pi_h(u_h - u)$,

$$\int_\Omega \eta \cdot \tau \, dx + \int_\Omega \mathrm{div}\, \tau w \, dx = 0 \qquad \text{for all } \tau \in H(\mathrm{div}, \Omega),$$
$$\int_\Omega \mathrm{div}\, \eta \, v \, dx = -\int_\Omega \Pi_h(u_h - u) v \, dx \quad \text{for all } v \in L^2(\Omega).$$

We recall that $\eta \in H^1[(\Omega)]^n$ thanks to elliptic regularity. Thus, the interpolation I_{RT} is well defined. We test the second equation with $-(u_h - \Pi_h u)$ and obtain from the commuting diagram property of the interpolation I_{RT} and the solution properties of u and u_h that that

$$\begin{aligned}\|\Pi_h(u_h - u)\|^2_{L^2(\Omega)} &= -\int_\Omega \operatorname{div}\eta\, \Pi_h(u_h - u)\,dx \\ &= -\int_\Omega \operatorname{div} I_{RT}\eta\,(u_h - u)\,dx = \int_\Omega (\sigma_h - \sigma) I_{RT}\eta\,dx.\end{aligned}$$

We add and subtract η and use the first equation for η, which leads to

$$\int_\Omega (\sigma_h - \sigma) I_{RT}\eta\,dx = \int_\Omega (\sigma_h - \sigma)(I_{RT}\eta - \eta)\,dx - \int_\Omega \operatorname{div}(\sigma_h - \sigma) w\,dx.$$

The Galerkin equations for σ_h show that $\operatorname{div}(\sigma_h - \sigma)$ is L^2 orthogonal to the piecewise constant functions, so that we can subtract $\Pi_h w$ from w in the last integral. Thus, combining the previous two chains of identities with the Cauchy inequality yields

$$\begin{aligned}&\|\Pi_h(u_h - u)\|^2_{L^2(\Omega)} \\ &\quad\le \|\sigma - \sigma_h\|_{L^2(\Omega)}\|\eta - I_{RT}\eta\|_{L^2(\Omega)} + \|\operatorname{div}(\sigma - \sigma_h)\|_{L^2(\Omega)}\|w - \Pi_h w\|_{L^2(\Omega)}.\end{aligned}$$

Due to the H^2 regularity and $\eta = \nabla w$, we can use the error estimate for I_{RT} and the piecewise Poincaré inequality for $w - \Pi_h w$ and the elliptic regularity estimate to obtain

$$\|\eta - I_{RT}\eta\|_{L^2(\Omega)} + \|w - \Pi_h w\|_{L^2(\Omega)} \lesssim h\|w\|_{H^2(\Omega)} \lesssim h\|\Pi(u - u_h)\|_{L^2(\Omega)}.$$

The assertion follows from combining the foregoing two estimates. □

Corollary 2.40. *We have*

$$\|u - u_h\|_{H^{-1}(\Omega)} \lesssim h(\|\sigma - \sigma_h\|_{H(\operatorname{div},\Omega)} + \|u - u_h\|_{L^2(\Omega)}).$$

Proof. This follows from adding and subtracting $\Pi_h u$, the triangle inequality, direct computations with the H^{-1} norm and the projection Π_h, and the Poincaré inequality. □

2.3.6. Estimates based on the hypercircle identity. The following result can be found in the literature under the name *Prager–Synge theorem* or *hypercircle identity.*

Lemma 2.41. *Let $\Omega \subseteq \mathbb{R}^n$ be an open and bounded Lipschitz domain and let $u \in H^1_0(\Omega)$ solve Poisson's problem with right-hand side $f \in L^2(\Omega)$. Then, any $\sigma \in H(\operatorname{div},\Omega)$ with $-\operatorname{div}\sigma = f$ and any $v \in H^1_0(\Omega)$ satisfy the hypercircle identity*

$$\|\nabla(u - v)\|^2_{L^2(\Omega)} + \|\nabla u - \sigma\|^2_{L^2(\Omega)} = \|\nabla v - \sigma\|^2_{L^2(\Omega)}.$$

Proof. In the norm on the right-hand side we add and subtract ∇u and apply the binomial theorem to the squared norm. The result is the left-hand side plus the mixed expression

$$-2\int_\Omega \nabla(u - v)\cdot(\nabla u - \sigma)\,dx,$$

which equals zero as can be seen by integration by parts because $\operatorname{div}(\nabla u - \sigma) = 0$. □

The fact that the right-hand side in the the hypercircle identity is independent of the unknown function u makes the result very useful for *a posteriori error estimation.*

If we choose $v = u_h \in S^1_0(\mathcal{T})$ to be the Galerkin approximation to u with the standard finite element method, the hypercircle identity implies the error bound

$$\|\nabla(u-u_h)\|_{L^2(\Omega)} \le \|\nabla u_h - \sigma\|_{L^2(\Omega)} \quad \text{for any } \sigma \in H(\operatorname{div},\Omega) \text{ with } -\operatorname{div}\sigma = f.$$

Once we make a choice for σ, the right-hand side is fully computable and is a guaranteed bound (there are no constants is the estimate) to the Galerkin error. Vice versa, if f is piecewise constant and $\sigma_h \in RT_0(\mathcal{T})$ is the discrete solution by the Raviart–Thomas method, we have the estimate

$$\|\nabla u - \sigma_h\|_{L^2(\Omega)} \le \|\nabla v - \sigma\|_{L^2(\Omega)} \quad \text{for any } v \in H^1_0(\Omega).$$

Such bounds are called *a posteriori* error estimates because they involve information of the discrete solution and thus are evaluated after the computation.
A direct consequence is:

COROLLARY 2.42. *Let $\Omega \subseteq \mathbb{R}^n$ be an open and bounded Lipschitz polytope triangulated by $\mathcal{T}$, let $f \in P_0(\mathcal{T})$ and let $u_h \in S^1_0(\mathcal{T})$ and $\sigma_h \in RT_0(\mathcal{T})$ be the approximations to $u \in H^1_0(\Omega)$ resp. ∇u by the standard resp. Raviart–Thomas FEM, where u solves Poisson's equation $-\Delta u = f$. We have the guaranteed a posteriori error bound*

$$\|\nabla(u-u_h)\|^2_{L^2(\Omega)} + \|\nabla u - \sigma_h\|^2_{L^2(\Omega)} = \|\nabla u_h - \sigma_h\|^2_{L^2(\Omega)}.$$

The following result states that the standard FEM and the Raviart–Thomas FEM are in some sense the optimal choice. We denote

$$Q_h(f) = \{\tau_h \in RT_0(\mathcal{T}) : -\operatorname{div}\tau_h = f\}.$$

LEMMA 2.43. *Under the conditions of Corollary 2.42 we have*

$$\|\nabla u_h - \sigma_h\|_{L^2(\Omega)} = \min_{v_h \in S^1_0(\mathcal{T})} \min_{\tau_h \in Q_h(f)} \|\nabla v_h - \tau_h\|_{L^2(\Omega)}.$$

PROOF. For any $\tau_h \in Q_h(f)$ and any $v_h \in S^1_0(\mathcal{T})$, the hypercircle reads

$$\|\nabla(u-v_h)\|^2_{L^2(\Omega)} + \|\nabla u - \tau_h\|^2_{L^2(\Omega)} = \|\nabla v_h - \tau_h\|_{L^2(\Omega)}.$$

Since u_h is the best approximation in the energy norm, the left-hand side is minimal for $v_h = u_h$, and therefore we have shown

$$\|\nabla u_h - \tau_h\|_{L^2(\Omega)} = \min_{v_h \in S^1_0(\mathcal{T})} \|\nabla v_h - \tau_h\|_{L^2(\Omega)}.$$

It remains to show that this expression is minimal for $\tau_h = \sigma_h$. To this end, we minimize the left-hand side over $Q_h(f)$, which is equivalent to

$$\frac{1}{2}\|\tau_h\|^2_{L^2(\Omega)} - \int_\Omega \nabla u_h \cdot \tau_h \, dx \to \min.$$

The Euler–Lagrange equation for the minimizer $\xi_h \in Q_h(f)$ of this quadratic minimization problem is (after integration by parts)

$$\int_\Omega \xi_h \cdot \tau_h \, dx = \int_\Omega \nabla u_h \cdot \tau_h \, dx = 0 \qquad \text{for all } \tau_h \in Q_h(0).$$

The inf-sup condition for the Raviart–Thomas method shows that there exists a Lagrange multiplier $w_h \in P_0(\mathcal{T})$ such that

$$\begin{aligned} \int_\Omega \xi_h \cdot \tau_h \, dx + \int_\Omega w_h \operatorname{div}\tau_h \, dx &= 0 && \text{for all } \tau_h \in RT_0(\mathcal{T}) \\ \int_\Omega \operatorname{div}\xi_h v_h \, dx &= -\int_\Omega f v_h \, dx && \text{for all } v_h \in P_0(\mathcal{T}). \end{aligned}$$

This shows that $\xi_h = \sigma_h$ is the solution to the Raviart–Thomas system. This establishes the asserted identity. □

The foregoing result has shown that for bounding the error in the standard FEM, the optimal choice from $RT_0(\mathcal{T})$ for the upper bound is the result of the Raviart–Thomas FEM; and that the best choice from $S^1_0(\mathcal{T})$ for bounding the Raviart–Thomas error is the solution to the standard FEM. The lemma has shown that this choice is sharp in the sense that the upper bound is bounded by the errors of the two methods. The disadvantage is that, for example, for bounding the error of the standard FEM, an additional mixed linear system of more or less the same size needs to be solved, which is considered too expensive. Instead, a suitable $\tau_h \in Q_h(f)$ can be designed by a local construction. We restrict our attention to $n = 2$ for simplicity. We observe that ∇u_h is piecewise divergence-free but not globally in $H(\operatorname{div}, \Omega)$. The jump of a (possibly vector-valued) function v across an edge E is denoted by $[v]_E := v|_{T_+} - v|_{T_-}$ for the two elements $T_\pm$ sharing E. For boundary faces there is only one element T_+ and we set $[v]_E := v|_{T_+}$. In every element we have $\nabla u_h|_T \in RT_0(T)$. Once we have designed a piecewise RT_0 function τ_h^{pw} (not in $H(\operatorname{div}, \Omega)$ in general) with the property that $-\operatorname{div} \tau_h^{\mathrm{pw}}|_T = f|_T$ on every $T \in \mathcal{T}$ and $[\tau_h^{\mathrm{pw}}]_E \cdot \nu_E = -[\nabla u_h]_E \cdot \nu_E$ for every interior edge E, we have that $\tau_h^{\mathrm{pw}} = \tau_h - \nabla u_h$ for an element $\tau_h \in Q(f)$. It remains to evaluate the norm of τ_h^{pw}.

A possible construction is as follows. For a vertex z of the triangulation $\mathcal{T}$, we recall the vertex patch ω_z, which is the interior of the union of all triangles containing z. We define the set $\mathcal{E}(z)$ of edges containing z and denote by φ_z the corresponding $S^1(\mathcal{T})$ nodal basis function. We design a piecewise RT_0 function τ_h^z supported on ω_z as follows. For every edge $E \notin \mathcal{E}(z)$ we set the degree of freedom $\int_E \tau_z \cdot \nu_E \, ds = 0$. The remaining degrees of freedom are related to two faces per triangle. They are fixed by the conditions

$$\int_{\partial T} \tau_h^z \cdot \nu_T \, ds = -\int_T f \varphi_z \, dx \quad \text{for every } T \subseteq \overline{\omega}_z$$
$$[\tau_h^z]_E \cdot \nu_E = -\frac{1}{2}[\nabla u_h]_E \cdot \nu_E \quad \text{for every } E \in \mathcal{E}(z).$$

If z is an interior vertex, a simple degree-of-freedom count reveals that such choice can be achieved. If z is a boundary vertex, we enforce the jump condition only on the interior edges. Recall that for boundary faces there is no condition on the normal trace for a piecewise polynomial field to belong to $H(\operatorname{div}, \Omega)$. We then obtain as many conditions as degrees of freedom if we consider the connectivity components of $\partial\Omega$. A practical implementation is outlined in [**Bra07**, III§9].

LEMMA 2.44. *The function $\tau_h^{\mathrm{pw}} := \sum_{z \in \mathcal{N}} \tau_h^z$ satisfies $-\operatorname{div} \tau_h^{\mathrm{pw}}|_T = f|_T$ on every $T \in \mathcal{T}$ and $[\tau_h^{\mathrm{pw}} \nu_E]_E \cdot = -[\nabla u_h]_E \cdot \nu_E$ for every interior edge E.*

PROOF. Since the nodal basis functions φ_z form a partition of unity, the design of the functions τ_h^z implies that

$$\sum_{z \in \mathcal{N}(T)} \int_T \operatorname{div} \tau_h^z \, dx = \sum_{z \in \mathcal{N}(T)} \int_{\partial T} \tau_h^z \cdot \nu_T \, ds = \int_T f \, dx$$

and therefore $-\operatorname{div} \tau_h^{\mathrm{pw}}|_T = f|_T$ on every $T \in \mathcal{T}$. Furthermore, any edge E is shared by two vertices z_1, z_2, such that

$$[\tau_h^{\mathrm{pw}}]_E \cdot \nu_E = [\tau_h^{z_1}]_E \cdot \nu_E + [\tau_h^{z_2}]_E \cdot \nu_E = -[\nabla u_h]_E \cdot \nu_E.$$

□

We conclude the following reliability estimate:

THEOREM 2.45. *Under the above assumptions (in particular f piecewise constant) we have the a posteriori error bound*

$$\|\nabla(u-u_h)\|_{L^2(\Omega)} \le \|\tau_h^{\mathrm{pw}}\|_{L^2(\Omega)}.$$

Algorithmic details on the implementation can be found in [**Bra07**, Chapter III §9].

REMARK 2.46. One can prove that the bound is also efficient, that is the converse estimate holds up to a constant,

$$\|\tau_h^{\mathrm{pw}}\|_{L^2(\Omega)} \lesssim \|\nabla(u-u_h)\|_{L^2(\Omega)}.$$

♦

REMARK 2.47. If $f \in L^2(\Omega)$ is not piecewise constant, we have

$$\|\nabla(u-u_h)\|_{L^2(\Omega)} \le \|\tau_h^{\mathrm{pw}}\|_{L^2(\Omega)} + \sqrt{\sum_{T\in\mathcal{T}} \frac{h_T^2}{\pi^2}\|f - \fint_T f\,dx\|^2_{L^2(T)}},$$

see Exercise 2.35. The additional term is referred to as *data oscillation.* ♦

2.4. Nonconforming FEM

2.4.1. The Crouzeix–Raviart element. For standard methods we assumed the conformity property $V_h \subseteq V$, which led to a convenient error analysis via Céa's lemma. The idea of nonconforming methods is to gain more flexibility (in whatever sense) of the discretization by giving up that constraint. In general we will therefore work with discrete space $V_h \not\subseteq V$. We start with the Crouzeix–Raviart element as a basic example. For simplicity we shall work in $\mathbb{R}^2$. As usual, $P_1(\mathcal{T})$ is the space of piecewise affine (but possibly discontinuous) functions. Given a triangulation $\mathcal{T}$ of our usual bounded, open, polygonal Lipschitz domain Ω, we define

$$CR^1(\mathcal{T}) := \{v \in P_1(\mathcal{T}) : v \text{ is continuous is the midpoints of interior faces}\}.$$

The version with homogeneous boundary conditions reads

$$CR^1_0(\mathcal{T}) := \{v \in CR^1(\mathcal{T}) : v \text{ vanishes in the midpoints of boundary faces}\}.$$

We want to use this space to approximate the Dirichlet problem for the Laplacian, but we have the obvious difficulty that $CR^1_0(\mathcal{T})$ is not a subspace of $H^1_0(\Omega)$. For piecewise regular objects such as $v_h \in CR^1(\mathcal{T})$, we can evaluate a piecewise gradient

$$\nabla_h v_h \in L^2(\Omega) \text{ defined by } (\nabla_h v_h)|_T = \nabla(v_h|_T) \text{ for any } T \in \mathcal{T}$$

and define

$$|\!|\!| v |\!|\!|_h := \|\nabla_h v\|_{L^2(\Omega)} \quad \text{for any piecewise } H^1\text{-regular function.}$$

We can show:

LEMMA 2.48. *The seminorm $|\!|\!| v |\!|\!|_h$ is a norm on the sum space $H^1_0(\Omega) + CR^1_0(\mathcal{T})$.*

PROOF. Exercise 2.39. □

The seminorm is induced by the bilinear form

$$a_h(v,w) := \int_\Omega \nabla_h v \cdot \nabla_h w\,dx \quad \text{for any } v, w \in H^1(\Omega) + CR^1(\mathcal{T}).$$

We have shown that a_h is an inner product on $CR^1_0(\mathcal{T})$, from which it is clear that, given $f \in L^2(\Omega)$, there exists a unique solution $u_h \in CR^1_0(\mathcal{T})$ to

$$a_h(u_h, v_h) = \int_\Omega f v_h\,dx \quad \text{for all } v_h \in CR^1_0(\mathcal{T}).$$

This is the the Crouzeix–Raviart (or nonconforming P_1) method for the Dirichlet problem of the Laplacian. For the implementation, we use the face-oriented basis functions with the property

$$\fint_F \psi_E \, ds = \delta_{E,F}$$

for interior faces E, F. We note that for piecewise affine functions stating that a function is continuous in a face midpoint is equivalent with the property that the average $\fint_E \cdot \, ds$ coincides on both neighbouring elements T_+ and T_-. On an element T with barycentric coordinates ϕ_1, ϕ_2, ϕ_3, and faces E_1, E_2, E_3 we use the convention that $\phi_j|_{E_j} = 0$, that is E_j is opposite to the vertex z_j. The local basis function ψ_{E_j} then reads

$$\psi_{E_j} = 1 - 2\varphi_j.$$

It is direct to verify that therefore the local stiffness matrix equals four times the local stiffness matrix of the standard FEM.
The nonconforming interpolation operator is defined via

$$I_h v := \sum_{E \in \mathcal{E}} \fint_E v \, ds \psi_E \quad \text{for any } v \in H^1(\Omega) + CR^1(\mathcal{T}).$$

It has the following important property.

Lemma 2.49 (projection property). *The nonconforming interpolation satisfies for any $v \in H^1(\Omega)$*

$$\nabla_h I_h v = \Pi_0 \nabla v.$$

That is, the piecewise gradient of the interpolated function equals the best approximation of the gradient by piecewise constants.

Proof. Exercise 2.38. □

We proceed with a basic error estimate.

Theorem 2.50. *Let Ω be an open and connected polygonal Lipschitz domain and assume that the solution u to the Poisson problem with $f \in L^2(\Omega)$ satisfies $u \in H^1_0(\Omega) \cap H^2(\Omega)$. Then*

$$|\!|\!| u - u_h |\!|\!|_h \lesssim h \|D^2 u\|_{L^2(\Omega)}.$$

Proof. We write $w_h := I_h u - u_h$ and use the triangle inequality

$$|\!|\!| u - u_h |\!|\!|_h \le |\!|\!| u - I_h u |\!|\!|_h + |\!|\!| w_h |\!|\!|_h$$

and observe that the square of the second term on the right-hand side satisfies

$$|\!|\!| I_h u - u_h |\!|\!|_h^2 = a_h(I_h u - u_h, w_h) = a_h(I_h u, w_h) - \int_\Omega f w_h \, dx$$

because w_h belongs to the finite element space. We use the projection property of I_h and integration by parts for the term including $I_h u$ and compute

$$a_h(I_h u, w_h) = a_h(u, w_h) = \int_\Omega f w_h \, dx - \sum_{E \in \mathcal{E}} \int_E \nabla u \cdot \nu_E [w_h]_E \, ds$$

where $[\cdot]_E$ denotes as usual the jump across E (for boundary faces, we define it as the usual trace) and where we have used that $\nabla u \cdot \nu_E$ does not jump; indeed ∇u is H^1 regular. On any interior face E, the jump $[w_h]_E$ has vanishing average and is thus orthogonal to any constant function. We compute

$$\begin{aligned} \int_E \nabla u \cdot \nu_E [w_h]_E \, ds &= \int_E \nabla_h (u - I_h u) \cdot \nu_E [w_h]_E \, ds \\ &\le \|\nabla_h (u - I_h u)|_T\|_{L^2(E)} \|[w_h - \fint_E w_h \, ds]\|_{L^2(E)}. \end{aligned}$$

With triangle, trace, and Poincaré inequalities as well as Exercise 2.40, we deduce

$$\int_E \nabla u \cdot \nu_E [w_h]_E \, ds \lesssim h \|D^2 u\|_{L^2(T_+ \cup T_-)} \|\nabla_h w_h\|_{L^2(T_+ \cup T_-)}.$$

Altogether, we conclude the stated result from the finite overlap of face patches and the combination with the above arguments. □

REMARK 2.51. In the previous proof we could not use Céa's lemma. Instead, we directly worked with the H^2 regularity of the solution. This assumption can be relaxed with a more elaborate proof. ♦

2.4.2. Application to the Stokes equations. We recall the Stokes equations. Given $f \in L^2(\Omega)$, we seek $u \in [H^1_0(\Omega)]^2$ and $p \in L^2_0(\Omega)$ such that

$$\begin{aligned} -\Delta u + \nabla p &= f \quad \text{in } [H^{-1}(\Omega)]^2 \\ \operatorname{div} u &= 0 \quad \text{in } L^2_0(\Omega). \end{aligned}$$

Here, u is a vector field and Δ is defined component-wise. As usual, $L^2_0(\Omega)$ are the L^2 functions with vanishing integral over Ω. Since $\int_\Omega \operatorname{div} u \, dx = 0$ due to integration by parts, the second equation is indeed valid pointwise almost everywhere. The problem can be put in a saddle-point formulation. We set $V = [H^1_0(\Omega)]^2$, $M := L^2_0(\Omega)$ and

$$a(v,w) = \int_\Omega Dv : Dw \, dx, \quad b(v,q) = -\int_\Omega q \operatorname{div} v \, dx, \quad F(v) = \int_\Omega f \cdot v \, dx, \quad G = 0$$

and see that the above equation is equivalent to the usual saddle-point problem with this specific choices. The problem admits a unique solution. The proof obviously requires an inf-sup condition for the form b. We quote the result, which we will not prove in this lecture.

THEOREM 2.52. *Given an open, bounded, connected Lipschitz domain Ω, there exists β such that*

$$0 < \beta = \inf_{q \in L^2_0(\Omega) \setminus \{0\}} \sup_{v \in [H^1_0(\Omega)]^2 \setminus \{0\}} \frac{\int_\Omega q \operatorname{div} v \, dx}{\|Dv\|_{L^2(\Omega)} \|q\|_{L^2(\Omega)}}$$

for some β.

We denote by Z the subspace of V of divergence-free vector fields

$$Z := \{v \in V : \operatorname{div} v = 0\}.$$

The solution u from the Stokes equations belongs to Z satisfies

$$a(u,v) = F(v) \quad \text{for all } v \in Z.$$

It is known from previous lectures that the design of Galerkin methods in Z is very difficult, see Exercise 2.11. Discretizing the saddle–point problem is easier, but the resulting approximation will not be pointwise divergence-free in general. The advantage of a nonconforming discretization is that the discrete velocity field u_h is piecewise divergence-free, at the expense of the nonconformity $u_h \notin V$. We denote $V_h = [CR^1_0(\mathcal{T})]^2$, $M_h := P_0(\mathcal{T}) \cap L^2_0(\Omega)$ and

$$a_h(v,w) = \int_\Omega D_h v : D_h w \, dx, \quad b_h(v,q) = -\int_\Omega q \operatorname{div}_h v \, dx.$$

The nonconforming method seeks $u_h \in V_h$ and $p_h \in M_h$ such that

$$\begin{aligned} a_h(u_h, v_h) + b_h(v_h, p_h) &= F(v_h) && \text{for all } v_h \in V_h \\ b_h(u_h, q_h) &= 0 && \text{for all } q_h \in M_h. \end{aligned}$$

LEMMA 2.53. *The discrete Stokes system has a unique solution (u_h, p_h).*

PROOF. It suffices to check the discrete inf-sup condition. Given $q_h \in M_h$, the continuous inf-sup condition and the projection property of I_h show that

$$0 < \beta = \sup_{v \in [H_0^1(\Omega)]^2 \setminus \{0\}} \frac{\int_\Omega q_h \operatorname{div} v \, dx}{\|Dv\|_{L^2(\Omega)} \|q_h\|_{L^2(\Omega)}} \sup_{v \in [H_0^1(\Omega))]^2 \setminus \{0\}} \frac{\int_\Omega q_h \operatorname{div}_h I_h v \, dx}{\|Dv\|_{L^2(\Omega)} \|q_h\|_{L^2(\Omega)}}.$$

The projection property furthermore implies $\|D_h I_h v\|_{L^2(\Omega)} \leq \|Dv\|_{L^2(\Omega)}$. This implies the discrete inf-sup condition. □

It is not difficult to see that a solution u_h will satisfy $\operatorname{div}_h u_h = 0$. We consider $Z_h := \{v_h \in V_h : \operatorname{div}_h v_h = 0\}$.

LEMMA 2.54. *The discrete solution u_h to the nonconforming Stokes discretization satisfies $u_h \in Z_h$ and*

$$a_h(u_h, v_h) = F(v_h) \quad \textit{for all } v_h \in Z_h.$$

PROOF. This follows from testing with elements from Z_h. □

It is not difficult to obtain a basic a priori error estimate.

THEOREM 2.55. *Assume the solution pair (u, p) to the Stokes system with $f \in L^2(\Omega)$ satisfies $u \in [H_0^1(\Omega)] \cap [H^2(\Omega)]^2$ and $p \in L_0^2(\Omega) \cap H^1(\Omega)$. Then, the error of the nonconforming FEM discretization satisfies*

$$|||u - u_h|||_h + \|p - p_h\|_{L^2(\Omega)} \lesssim h(\|D^2 u\|_{L^2(\Omega)} + \|\nabla p\|_{L^2(\Omega)}).$$

REMARK 2.56. These regularity assumptions are satisfied on convex domains (PDE literature). ♦

PROOF OF THEOREM 2.55. It suffices to bound the norms of the errors $I_h u - u_h$ and $\Pi_0 p - p_h$ (use the triangle inequality and known bounds). The discrete inf-sup condition states

$$\begin{aligned} &|||I_h u - u_h|||_h + \|\Pi_0 p - p_h\|_{L^2(\Omega)} \\ &\quad \lesssim \sup_{\substack{|||w_h|||_h = 1 \\ \|q_h\|_{L^2(\Omega)} = 1}} \left[a_h(I_h u - u_h, w_h) + b_h(w_h, \Pi_0 p - p_h) + b_h(I_h u - u_h, q_h) \right]. \end{aligned}$$

The projection properties of I_h and Π_0 and the constraint on the divergence show that the last term on the right-hand side equals zero and that

$$a_h(I_h u - u_h, w_h) + b_h(w_h, \Pi_0 p - p_h) = a_h(u, w_h) + b_h(w_h, p) - \int_\Omega f \cdot w_h \, dx.$$

We proceed in a similar fashion as in the convergence proof for the Poisson equation. From piecewise integration by parts we obtain

$$a_h(u, w_h) + b_h(w_h, p) = \int_\Omega f w_h \, dx - \sum_{E \in \mathcal{E}} \int_E ((Du - pI_{2\times 2})\nu_E) \cdot [w_h]_E \, ds.$$

The conclusion of the proof is similar as in the Poisson case and left as an exercise. □

We have seen that the nonconforming method directly produces piecewise divergence-free solutions. It is possible to design a local basis of Z_h in an explicit construction:

- For each interior edge E we take a function $\alpha_E \in CR_0^1(\mathcal{T})$ such that

  $$⨍_E \alpha_E \cdot \nu_E \, ds = 0, \quad ⨍_E \alpha_E \cdot t_E \, ds = 1, \quad ⨍_F \alpha_E \, ds = 0 \text{ for } F \neq E.$$

  Here $t_E = (-\nu_{E,2}, \nu_{E,1})$ is a unit tangent vector.

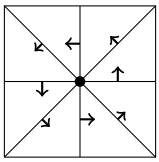


FIGURE 2. Orientation of the normal vectors $\hat{\nu}_E$ around the vertex z

- For each interior vertex z with set of edges $\mathcal{E}(z)$ containing z we define $\alpha_z \in CR_0^1(\mathcal{T})$ as follows. All tangential components are set to zero. Also, the normal components are set to zero on those edges that do not touch z,

$$\fint_E \alpha_z \cdot t_E \, ds = 0 \quad \text{for all } E \in \mathcal{E}(\Omega) \quad \text{and} \quad \fint_E \alpha_z \cdot \nu_E \, ds = 0 \quad \text{for all } E \notin \mathcal{E}(z).$$

  For any edge E touching z we choose a normal vector $\hat{\nu}_E$ with counterclockwise orientation (see Figure 2 and choose

$$\int_E \alpha_z \cdot \hat{\nu}_E \, ds = 1.$$

It is not difficult to check that α_z and α_E belong to Z_h and are linear independent. By a dimension argument (see Exercise 2.10) it can then be shown that the functions form a basis of Z_h if the domain is simply connected. Details are worked out in Exercise 2.43.

2.4.3. Morley element. We consider a variational problem in the space $H_0^2(\Omega)$, the biharmonic problem. Given $f \in L^2(\Omega)$ for simplicity, it seeks a function u such that

$$\Delta^2 u = f \text{ in } \Omega \quad \text{and} \quad u = \partial u/\partial \nu = 0 \text{ on } \partial\Omega.$$

It is easy to calculate via integration by parts that a sufficiently smooth function $u \in H_0^2(\Omega)$ satisfies

$$\int_\Omega \Delta^2 u \varphi \, dx = \int_\Omega D^2 u : D^2 \varphi \, dx = \int_\Omega \Delta u \Delta \varphi \, dx \quad \text{for } \varphi \in C_c^\infty(\Omega).$$

The corresponding variational equality

$$\int_\Omega D^2 u : D^2 v \, dx = \int_\Omega f v \, dx \quad \text{for all } v \in H_0^2(\Omega)$$

has a unique solution by the Riesz representation theorem in $H_0^2(\Omega)$. If V_h is a subspace of $H_0^2(\Omega)$, the Galerkin projection is easily defined and standard theory can be used to establish an a priori error analysis. However, it turns out that the construction of H^2 conforming piecewise polynomial finite element spaces is rather complicated. The three simplest choices are the Argyris element, the Hsieh–Clough–Tocher (HCT) element, or the Bogner–Fox–Schmid (BFS) element from Figure 3.

We will use a nonconforming element that allows a much simpler local construction by giving up certain continuity constraints. The Morley element is the following (formal) finite element for a triangle T

$$(T, P_2(T), \{\delta_z, \fint_E \frac{\partial \bullet}{\partial \nu_T} \, ds : z \in \mathcal{N}(T), E \in \mathcal{E}(T)\}),$$

that is, the shape function are the quadratic polynomials and the degrees of freedom are the point evaluations at the three vertices and the evaluations of the averages

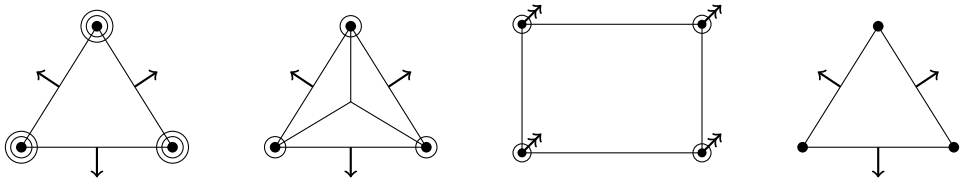


FIGURE 3. Mnemonic diagrams of some finite elements for the biharmonic equation: Argyris, HCT, BFS (definitions see [**Cia78**]), and Morley.

of the normal derivative over the three edges of the triangle. The Morley finite element space is

$$M_0(\mathcal{T}) := \left\{ v \in P_2(\mathcal{T}) : \begin{array}{l} v \text{ continuous at the interior vertices,} \\ v = 0 \text{ at boundary vertices} \\ \partial v/\partial \nu_E \text{ continuous at the interior edges' midpoints,} \\ \partial v/\partial \nu = 0 \text{ at boundary edges' midpoints} \end{array} \right\}.$$

Given $f \in L^2(\Omega)$, the discrete problem seeks $u_h \in M_0(\mathcal{T})$ such that

$$\int_\Omega D_h^2 u_h : D_h^2 v_h \, dx = \int_\Omega f v_h \, dx \quad \text{for all } v_h \in M_0(\mathcal{T}).$$

It is easy to check that the left hand side defines a positive definite bilinear form: if the piecewise Hessian $D_h^2 v_h$ of v_h is zero, then v_h must be piecewise affine. The continuity at interior vertices implies then that v_h is continuous and thus in $S^1(\mathcal{T})$. The continuity of the normal derivatives over the edge midpoints shows that v_h must be globally affine and, by the boundary conditions imposed on $M_0(\mathcal{T})$, therefore is the zero function. Hence, there exists a unique solution u_h to the discrete problem. The main tool in the error analysis is again a nonconforming interpolation operator, which is defined via the degrees of freedom. Given $v \in H_0^2(\Omega)$, the element $I_h^M v \in M_0(\mathcal{T})$ is uniquely defined by the conditions

$$(v - I_h^M v)(z) = 0 \text{ for all } z \in \mathcal{N} \quad \text{and} \quad \fint_E \frac{\partial(v - I_h v)}{\partial \nu_E}(z) = 0 \text{ for all } E \in \mathcal{E}.$$

With arguments similar to those for the Crouzeix–Raviart element we show the projection property for the Hessian

$$D_h^2 I_h^M v = \Pi_0 D^2 v.$$

There is indeed a close connection between the Morley and the Crouzeix–Raviart method. First, it is directly verified that

$$\nabla_h M_0(\mathcal{T}) \subseteq CR_0^1(\mathcal{T}) \quad \text{and} \quad I_h^{CR} \nabla v = \nabla_h I_h^M v.$$

We will now prove that the horizontal sequences in Figure 4 are exact and that the diagram commutes. We work with the operators

$$\operatorname{Curl} v = \begin{pmatrix} -\partial_y u \\ \partial_x u \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \nabla u \quad \text{and} \quad \operatorname{rot} \phi = \partial_x \phi_2 - \partial_y \phi 1$$

for scalar functions v and vector fields ϕ. The piecewise counterparts are as usual denoted with the index h.

THEOREM 2.57. *The diagram of Figure 4 commutes. If Ω is simply connected, the horizontal sequences in Figure 4 are exact and the diagram commutes.*

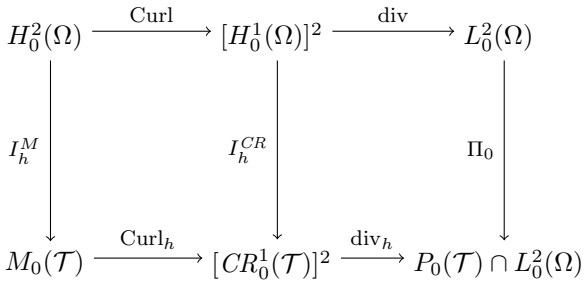


FIGURE 4. Curl-div complex.

PROOF. The commuting property is a direct consequence of the projection properties of the respective interpolation operators. It is a classical result that the first row is an exact sequence and we are left with showing this property for the second row. Clearly, $\mathrm{div}_h \mathrm{Curl}_h = 0$, which implies the complex property

$$\mathrm{Curl}_h M_0(\mathcal{T}) \subseteq Z_h$$

where

$$Z_h := \{v_h \in CR_0^1(\mathcal{T}) : \mathrm{div}_h v_h = 0\}.$$

For showing $\mathrm{Curl}_h M_0(\mathcal{T}) = Z_h$ it suffices to compare dimensions. We have previously shown that the dimension of Z_h equals $\mathrm{card}(N(\Omega)) + \mathrm{card}(\mathcal{E}(\Omega))$. This is precisely the number of degrees of freedom of the Morley element and thus the dimension of $M_0(\mathcal{T})$. The kernel of Curl_h, namely the piecewise constant functions, has only a trivial intersection with $M_0(\mathcal{T})$. □

From the above we observe that the solution u to the Stokes system with right-hand side f can be written as $\mathrm{Curl}\,\phi$ for some $\phi \in H_0^2(\Omega)$. We then have $-\Delta \mathrm{Curl}\,\phi + \nabla p = f$. Taking rot of the equation leads to

$$\Delta^2 \varphi = -\mathrm{rot}\, f$$

because $\mathrm{rot}\,\Delta\,\mathrm{Curl} = \Delta^2$. If the distribution $\mathrm{rot}\, f$ is an L^2 function, this can be directly discretized with the Morley element. Alternatively, we can discretize the right-hand side with the linear form

$$\int_\Omega f \cdot \mathrm{Curl}_h v_h \, dx \quad \text{for } v_h \in M_0(\mathcal{T}).$$

The resulting method with produce $u_h = \mathrm{Curl}_h \varphi_h$, which is the solution to the Crouzeix–Raviart method. In this sense, the structure from the continuous setting is preserved by the nonconforming spaces. In fluid mechanics, the function ϕ is called *stream function*.

2.4.4. The Helmholtz decomposition. The Helmholtz theorem is a classical result stating that any (unstructured) L^2 vector field can be decomposed as a gradient field and a divergence-free field. In what follows, we denote

$$\mathfrak{Z} = H(\mathrm{div}^0, \Omega) = \{\tau \in H(\mathrm{div}, \Omega) : \mathrm{div}\,\tau = 0\}.$$

LEMMA 2.58 (Helmholtz decomposition). *Let $\Omega \subseteq \mathbb{R}^n$ be an open, bounded, connected Lipschitz domain and let $p \in [L^2(\Omega)]^n$. Then there exist a unique $\alpha \in H_0^1(\Omega)$ and a unique $R \in \mathfrak{Z}$ such that*

$$p = \nabla \alpha + R.$$

The decomposition is $L^2(\Omega)$-orthogonal.

PROOF. Let $\alpha \in H^1_0(\Omega)$ denote the solution to $-\Delta\alpha = -\operatorname{div} p$ and set $R := p - \nabla\alpha$. Then we have $\operatorname{div} R = 0$ and thus the claimed decomposition. The orthogonality is easily checked with integration by parts, $\int_\Omega \nabla\alpha \cdot R\,dx = -\int_\Omega \alpha \operatorname{div} R\,dx = 0$. □

In shorthand notation, we write

$$[L^2(\Omega)]^n = \nabla H^1_0(\Omega) \oplus \mathfrak{Z}.$$

The gradient part $\nabla\alpha$ is sometimes called *Helmholtz projector* in the literature. A remarkable structure of the nonconforming method is that is satisfies a discrete analogue of the Helmholtz decomposition. Thereby, we also find a close connection to the Raviart–Thomas space. We will prove

$$[P_0(\mathcal{T})]^n = \nabla_h CR^1_0(\Omega) \oplus \mathfrak{Z}_h$$

where $\mathfrak{Z}_h := RT_0(\mathcal{T}) \cap \mathfrak{Z}$.

LEMMA 2.59 (discrete Helmholtz decomposition). *Let $\Omega \subseteq \mathbb{R}^n$ be an open, bounded, connected Lipschitz polytope and let $p_h \in [P_0(\mathcal{T})]^n$. Then there exist a unique $\alpha_h \in CR^1_0(\Omega)$ and a unique $R_h \in \mathfrak{Z}_h$ such that*

$$p_h = \nabla_h \alpha_h + R_h.$$

The decomposition is $L^2(\Omega)$-orthogonal.

PROOF. As in the continuous case, we denote by $\alpha_h \in CR^1_0(\mathcal{T})$ the unique solution to

$$\int_\Omega \nabla_h\alpha_h \cdot \nabla_h v_h\,dx = \int_\Omega p_h \cdot \nabla_h v_h\,dx \quad \text{for all } v_h \in CR^1_0(\mathcal{T})$$

and denote $R_h := p_h - \nabla_h\alpha_h$. Clearly, R_h is piecewise constant. We denote by ψ_F the Crouzeix–Raviart basis function with respect to the interior face $F \in \mathcal{F}(\Omega)$ (in 2d this is an interior edge). With this test function observe from the above solution property and integration by parts that

$$0 = \int_\Omega R_h \cdot \nabla_h\psi_F\,dx = \int_F [R_h]_F \cdot \nu_F \psi_F\,ds = [R_h]_F \cdot \nu_F \int_F \psi_F\,ds.$$

We conclude that R_h does not have normal jumps and therefore belongs to $H(\operatorname{div},\Omega)$. Hence, $R_h \in \mathcal{Z}_h$. The orthogonality of the decomposition follows from integration by parts:

$$\int_\Omega \nabla_h\alpha_h \cdot R_h\,dx = -\int_\Omega \alpha_h \operatorname{div} R_h\,dx + \sum_{F\in\mathcal{F}(\Omega)} R_h \cdot \nu_F [\alpha_h]_F\,ds = 0$$

because the jumps of α_h have vanishing integral mean over the faces. □

Instead of working with explicit gradients, we can equivalently work with the orthogonal complement of $\mathfrak{Z}$ for solving the Poisson equation. We denote by $\Gamma := \nabla H^1_0(\Omega)$ the space of gradients and observe

$$\Gamma = \mathfrak{Z}^\perp.$$

Given $f \in L^2(\Omega)$, assume we are given any vector field $\varphi \in [L^2(\Omega)]^n$ with $-\operatorname{div}\varphi = f$. Then, the Poisson equation $-\Delta u = f$ is equivalent to finding $\gamma \in \Gamma$ with

$$\int_\Omega \gamma \cdot \tau\,dx = \int_\Omega \varphi \cdot \tau\,dx \quad \text{for all } \tau \in \Gamma.$$

The constraint $\gamma \in \Gamma$ can be encoded with a multiplier $z \in \mathfrak{Z}$. The mixed problem is then to find $(\gamma, z) \in [L^2(\Omega)]^n \times \mathfrak{Z}$ such that

$$\int_\Omega \gamma \cdot \tau \, dx + \int_\Omega z \cdot \tau \, dx = \int_\Omega \varphi \cdot \tau \, dx \quad \text{for all } \tau \in [L^2(\Omega)]^n$$
$$\int_\Omega \gamma \cdot y \, dx = 0 \quad \text{for all } y \in \mathfrak{Z}.$$

For showing that this is indeed well-posed, we only need to check the inf-sup condition

$$0 < \beta = \inf_{y \in \mathfrak{Z} \setminus \{0\}} \sup_{\tau \in [L^2(\Omega)]^n \setminus \{0\}} \frac{\int_\Omega y \cdot \tau \, dx}{\|y\|_{H(\operatorname{div},\Omega)} \|\tau\|_{L^2(\Omega)}},$$

which is immediately verified (choose $\tau = y$).
On the discrete level, we can analogously write $\Gamma_h = \nabla_h CR_0^1(\mathcal{T})$ and

$$\Gamma_h = \mathfrak{Z}_h^\perp$$

where now the symbol $\perp$ indicates the orthogonal complement within $[P_0(\mathcal{T})]^n$. The discrete formulation of the above version of Poisson's equation is to find $\gamma_h \in \Gamma_h$ with

$$\int_\Omega \gamma_h \cdot \tau_h \, dx = \int_\Omega \varphi \cdot \tau_h \, dx \quad \text{for all } \tau_h \in \Gamma_h.$$

The mixed problem is then to find $(\gamma_h, z_h) \in [P_0(\mathcal{T})]^n \times \mathfrak{Z}_h$ such that

$$\int_\Omega \gamma_h \cdot \tau_h \, dx + \int_\Omega z_h \cdot \tau_h \, dx = \int_\Omega \varphi \cdot \tau_h \, dx \quad \text{for all } \tau_h \in [P_0(\mathcal{T})]^n$$
$$\int_\Omega \gamma_h \cdot y_h \, dx = 0 \quad \text{for all } y \in \mathfrak{Z}_h.$$

We note that this is a conforming method for the mixed problem (but of course $\Gamma_h \not\subseteq \Gamma$). This shows that the Crouzeix–Raviart method can be interpreted as a conforming method. For a particular choice of φ we can indeed recover the usual Crouzeix–Raviart solution such that $\nabla_h u_h = \gamma_h$.

LEMMA 2.60. *Let $f \in L^2(\Omega)$ be piecewise constant. If $\varphi \in RT_0(\mathcal{T})$ with $-\operatorname{div} \varphi = f$ is given as right-hand side in the above mixed problem, then $\nabla_h u_h = \gamma_h$.*

PROOF. We can decompose any discrete test function $\tau_h = \nabla_h \alpha_h + R_h$. We conclude from the orthogonality and the solution property of u_h that

$$\int_\Omega \nabla_h u_h \cdot \tau_h \, dx = \int_\Omega \nabla_h u_h \cdot \nabla_h \alpha_h \, dx = \int_\Omega f \alpha_h \, dx.$$

Since $f = -\operatorname{div} \varphi$ and φ is a Raviart–Thomas function, we can integrate by parts

$$\int_\Omega f \alpha_h \, dx = \int_\Omega \varphi \cdot \nabla_h \alpha_h \, dx = \int_\Omega \varphi \cdot \tau_h \, dx - \int_\Omega z_h \cdot \tau_h \, dx$$

where z_h is the orthogonal projection of φ onto $\mathfrak{Z}_h$. Therefore, $\nabla_h u_h$ solves the mixed problem with the multiplier z_h. □

COROLLARY 2.61. *Let f be piecewise constant. Let $\sigma_h \in RT_0(\mathcal{T})$ be the vector part of the mixed Raviart–Thomas solution and let u_h denote the Crouzeix–Raviart solution. Then*

$$\Pi_0 \sigma_h = \nabla_h u_h.$$

PROOF. It is easy to check that the L^2 projection of σ_h on $\mathfrak{Z}_h$ equals zero (first line of the mixed system) and that $-\operatorname{div} \sigma_h = f$ (second line of the mixed system). In the foregoing proof we have shown

$$\int_\Omega \nabla_h u_h \cdot \tau_h \, dx = \int_\Omega \sigma_h \cdot \tau_h \, dx \quad \text{for all } \tau_h \in [P_0(\mathcal{T})]^n,$$

which is equivalent to the asserted identity. □

2.A. Problems

PROBLEM 2.1. Let L be a linear and continuous map between Banach spaces X, Y. Prove $\ker(L^*) = L(X)^\circ$ and $\ker(L) = {}^\circ(L^*(Y^*))$.

PROBLEM 2.2. Let X, Y be Banach spaces and $L \in \mathcal{L}(X, Y)$. Prove that L is compact if and only if L^* is compact.
Hints (converse direction is similar):
1. An operator is called compact if it maps bounded sets to relatively compact sets.
2. Show that $A = \overline{L(B_1(0))}$ is compact if L is compact.
3. Given a bounded sequence in Y^*, show that it is uniformly bounded and equicontinuous over A. Show that there is a convergent subsequence in $C(A)$ (Arzelà–Ascoli).
4. Show that L^* maps that subsequence to a Cauchy sequence in X^*.

PROBLEM 2.3. For Hilbert spaces X, Y and a continuous linear map $L \in \mathcal{L}(X, Y)$, the map $L^H \in L(Y, X)$ defined by

$$\langle Lx, y\rangle_Y = \langle x, L^H y\rangle_X \quad \text{for any } x \in X, y \in Y$$

is called the *adjoint* of L. Prove

$$L^H = R_X^{-1} \circ L^* \circ R_Y$$

where R_X, R_Y denote the Riesz isometries of X, Y.

PROBLEM 2.4 (Lax–Milgram lemma). Let X be a real Hilbert space with inner product $\langle\cdot,\cdot\rangle_X$ and let $a : X \times X \to \mathbb{R}$ be a bilinear form satisfying the following two properties

- $\exists \beta > 0\, \forall (x, y) \in X^2 \quad |a(x, y)| \le \beta \|x\|_X \|y\|_Y$ (continuity)
- $\exists \alpha > 0\, \forall x \in X \quad \alpha \|x\|_X^2 \le a(x, x)$ (coercivity) .

Prove (using the Banach–Babuška–Nečas lemma) that there exists a unique map $T : X \to X$ with the property

$$a(x, y) = \langle Tx, y\rangle_X \quad \text{for all } (x, y) \in X^2.$$

The map T is linear, continuous, and invertible with

$$\|T\|_{L(X,X)} \le \beta \quad \text{and} \quad \|T^{-1}\|_{L(X,X)} \le \frac{1}{\alpha}.$$

PROBLEM 2.5 (computing with dual spaces). (a) Let $M \subseteq X$ be a subset of a Banach space X. Prove that M° is closed. *(Hint: Embedding in the bidual space.)*
(b) Let $L \in \mathcal{L}(X, Y)$ be a linear and continuous map between Banach spaces X, Y. Prove $\overline{L^*(Y^*)} = {}^\circ(\ker(L^{**}))$.
(c) Let X be reflexive and let $L \in \mathcal{L}(X, Y)$ be injective with $L(X) \subseteq Y$ dense. Prove that L^* is injective and $L^*(Y^*) \subseteq X^*$ is dense. *(Hint: separation theorem)*

PROBLEM 2.6. Prove that any closed subspace of a reflexive Banach space is reflexive.

PROBLEM 2.7 (computing with the orthogonal complement). Let X be a Hilbert space with the Riesz isomorphism $T : X \to X^*$; and let M be a reflexive Banach space.

(a) Prove that $Z^\perp = T^{-1}(Z^\circ)$ for any closed subspace $Z \subseteq X$.
(b) Let $B : X \to M^*$ be a linear map such that B^* has a bounded inverse on its range. Prove that $B : (\ker B)^\perp \to M^*$ is an isomorphism.

PROBLEM 2.8. Prove that the Stokes equations are a necessary condition for any minimizer of the constrained energy minimization problem. Show that, for sufficiently regular solutions, the Stokes equations can be written as $-\Delta u + \nabla p = f$ and $\operatorname{div} u = 0$. Here, Δ is the component-wise action of the Laplacian.

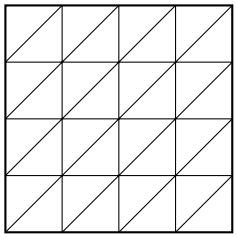
PROBLEM 2.9. Prove that div maps $[H^1_0(\Omega)]^2$ to a subspace of $L^2_0(\Omega)$.

PROBLEM 2.10 (Euler formulae). Let $\mathcal{T}$ be a triangulation of the simply-connected polygonal domain Ω. Prove

$$\operatorname{card}(\mathcal{T}) + \operatorname{card}(\mathcal{N}) = 1 + \operatorname{card}(\mathcal{E}) \quad \text{and} \quad 2\operatorname{card}(\mathcal{T}) + 1 = \operatorname{card}(\mathcal{N}) + \operatorname{card}(\mathcal{E}(\Omega)).$$

Here, as usual, $\mathcal{E}$ is the set of edges, $\mathcal{E}(\Omega)$ the set of interior edges, and $\mathcal{N}$ the set of vertices. What happens on planar domains with holes?

PROBLEM 2.11. *(conforming divergence-free functions are trivial)*
Let $\mathcal{T}$ be the criss triangulation of the unit square and let $u_h \in [S^1_0(\mathcal{T})]^2$ with $\operatorname{div} u_h = 0$. Prove that $u_h = 0$.
Hint: The criss triangulation is



.

PROBLEM 2.12. *(standard FEMs are unstable for Stokes)*
Let $\Omega = (0,1)^2$. Prove that the following discretizations of the Stokes equations lead to unstable saddle-point problems (i.e., the discrete inf-sup condition is violated): $V_h := [S^1_0(\mathcal{T})]^2$ and $M_h := P_0(\mathcal{T}) \cap L^2_0(\Omega)$ on the criss triangulation $\mathcal{T}_h$.
Hint: Use a dimension argument with the formulae from Problem 2.10.

PROBLEM 2.13. Prove Lemma 2.14.

PROBLEM 2.14. Prove Lemma 2.15.

PROBLEM 2.15. Prove the bound on $\|Du_h\|_{L^2(\Omega)}$ from the proof of Theorem 2.16.

PROBLEM 2.16. Implement the Mini finite element. As a test example, use the following data on the square $\Omega = (-1,1)^2$ (not the unit square): The right-hand side $f = 0$ is zero and the exact solution is

$$u(x_1, x_2) = \begin{pmatrix} 20x_1x_2^4 - 4x_1^5 \\ 20x_1^4x_2 - 4x_2^5. \end{pmatrix}$$

Choose the inhomogeneous Dirichlet data u_D according to u. Create convergence history plots for the error in the u variable.
Hint: An example of an implementation can be found on the course webpage.

PROBLEM 2.17. *(backward facing step)* Use the Mini element to simulate the flow over a backward facing step. Print the computed velocity and pressure and present the plots in the tutorial session. The parameters are:

- Domain: $\Omega = ((-2,8) \times (-1,1)) \setminus ([-2,0] \times [-1,0])$ (see Figure 5)
- Forcing term: $f = 0$,
- Dirichlet data: $u_D(x,y) = \begin{cases} (0,0) & \text{for } -2 < x < 8 \\ (-y(y-1)/10, 0) & \text{for } x = -2 \\ (-(y+1)(y-1)/80, 0) & \text{for } x = 8. \end{cases}$

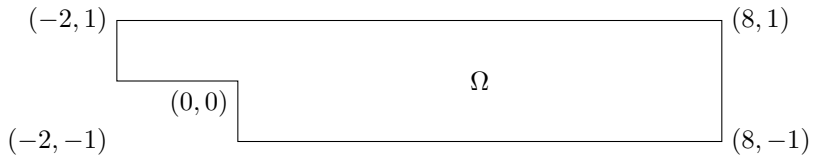


FIGURE 5. The backward facing step.

PROBLEM 2.18. Let $g \in H^{1/2}(\partial\Omega)$. Prove that the minimal extension, that is $u \in H^1(\Omega)$ with

$$\|\nabla u\|_{L^2(\Omega)} = \min_{\substack{v \in H^1(\Omega) \\ v|_{\partial\Omega} = g}} \|\nabla v\|_{L^2(\Omega)},$$

is given by the solution to

$$-\Delta u = 0 \quad \text{in } \Omega \quad \text{and} \quad u = g \quad \text{on } \partial\Omega.$$

PROBLEM 2.19 (negative Sobolev space). We know that $H_0^1(\Omega)$ is a Hilbert space when equipped with the inner product $\int_\Omega \nabla v \cdot \nabla w \, dx$. As such, it can be identified with its dual $H^{-1}(\Omega)$. We also know that $L^2(\Omega) \subseteq H^{-1}(\Omega)$. Does this imply that $L^2(\Omega)$ is also a subset of $H_0^1(\Omega)$? Give a complete explanation of this matter.

PROBLEM 2.20 (Gelfand triplet). Following the chain of the Gelfand triplet, we observe that, comparing with Y, a "smaller" space $X \subseteq Y$ will yield a "larger" dual space $Y^* \subseteq X^*$. If X is finite-dimensional $\dim(X) = n$, we know that also $\dim(X^*) = n$. Is therefore Y^* necessarily finite-dimensional?

PROBLEM 2.21. Let $\mathcal{T}$ be a regular triangulation of $\Omega \subseteq \mathbb{R}^n$ and let $v \in [P_1(\mathcal{T})]^n$ be a piecewise affine vector field. For each interior edge F with adjacent triangles T_+ and T_- (i.e., $F = T_+ \cap T_-$), the jump across F is defined by $[v]_F := v|_{T_+} - v|_{T_-}$. Prove that

$$v \in H(\operatorname{div}, \Omega) \iff [v \cdot \nu_F]_F = 0 \quad \text{for all interior edges } F$$

where ν_F is some normal vector of F.

PROBLEM 2.22. Prove that the normal trace is a surjective map from $\{v \in H(\operatorname{div}, \Omega) : \operatorname{div} v = 0\}$ to $\{g \in H^{-1/2}(\partial\Omega) : \langle g, 1 \rangle = 0\}$.

PROBLEM 2.23. Prove that the mixed form of the Poisson equation satisfies the properties of the Brezzi splitting theorem.

PROBLEM 2.24. Prove that the local basis functions $\psi_{T,E}$ satisfy the property (27).

PROBLEM 2.25. Write a routine (Python or pseudocode) that provides a global enumeration of all edges in a given mesh $\mathcal{T}$.

PROBLEM 2.26. Implement the mixed Raviart–Thomas method for the homogeneous Dirichlet problem of the Laplacian. Use the data from earlier exercises to compute experimental rates of convergence in different norms.

PROBLEM 2.27. Let T be a triangle. Prove that the following triplets $(T, \mathcal{P}, \mathcal{L})$ are finite elements in the sense of Ciarlet.

- The cubic Lagrange element: $\mathcal{P} = P_3(T)$ and $\mathcal{L}$ contains the point evaluations in the three vertices of T, in two interior points of each edge, and in the midpoint of T.
- The Crouzeix–Raviart element: $\mathcal{P} = P_1(T)$ and $\mathcal{L} := \{ \fint_E \cdot \, dx : E \in \mathcal{E}(T) \}$.

- The cubic Hermite element: $\mathcal{P} = P_3(T)$ and $\mathcal{L}$ contains the point evaluations in the three vertices and in the midpoint of T and the evaluation of the gradient in the vertices, that is

$$\mathcal{L} = \{v \mapsto v(z) : z \in \mathcal{N}(T)\} \cup \{v \mapsto \nabla v(z) : z \in \mathcal{N}(T)\} \cup \{v \mapsto v(\mathrm{mid}(T))\}.$$

- The Argyris element: $\mathcal{P} = P_5(T)$ and

$$\mathcal{L} = \{v \mapsto v(z), v \mapsto \nabla v(z), v \mapsto D^2 v(z) : z \in \mathcal{N}(T)\} \cup \{v \mapsto \fint_E \nabla v \cdot \nu_T \, ds : E \in \mathcal{E}(T)\}.$$

PROBLEM 2.28. Let T be a triangle and let $\mathcal{L}$ consist of the six functionals

$$\{\fint_E \cdot \, ds : E \in \mathcal{E}(T)\} \cup \{\fint_E \cdot s \, ds : E \in \mathcal{E}(T)\}$$

describing the first-order moments of a function over the three edges. Prove that $(T, P_2(T), \mathcal{L})$ is not a finite element in the sense of Ciarlet.

PROBLEM 2.29. Let $\mathcal{P}$ be an m-dimensional vector space and let $\mathcal{F}$ be a subset of $\mathcal{P}^*$ with m elements. Prove that the elements of $\mathcal{F}$ form a basis of $\mathcal{P}^*$ if and only if for any $v \in \mathcal{P}$ the relation $\langle F, v\rangle = 0$ for all $F \in \mathcal{F}$ implies $v = 0$.

PROBLEM 2.30. Prove that there exists a constant that only depends on the shape regularity such that

$$\|v - I_{RT} v\|_{L^2(T)} \le C h_T \|Dv\|_{L^2(T)} \quad \text{for any } v \in [H^1(T)]^2$$

and

$$\|\operatorname{div}(v - I_{RT} v)\|_{L^2(T)} \le C h_T \|\nabla \operatorname{div} v\|_{L^2(T)} \quad \text{for any } v \in [H^2(T)]^2.$$

PROBLEM 2.31. Prove that unit normal vectors transform as

$$\nu(x) = \frac{1}{|B^{-\top}\hat{\nu}(\hat{x})|} B^{-\top} \hat{\nu}(\hat{x}).$$

PROBLEM 2.32. Prove Lemma 2.38.

PROBLEM 2.33. Prove that the Raviart–Thomas interpolation is invariant under the Piola transform, i.e.,

$$I_{RT,\hat{T}} \hat{q} = \widehat{I_{RT,T} q}.$$

PROBLEM 2.34. Let $u_h \in S^1_0(\mathcal{T})$ be the standard FEM solution to the right-hand side $f \in L^2(\Omega)$. Let z be an interior vertex of $\mathcal{T}$ with hat function φ_z. Prove that

$$\frac{1}{2} \sum_{E \in \mathcal{E}(z)} \int_E [\nabla u_h]_E \cdot \nu_E \, ds = \int_{\omega_z} f \varphi_z \, dx$$

for the set $\mathcal{E}(z)$ of edges containing z.

PROBLEM 2.35. Prove that for $f \in L^2(\Omega)$ the following error bound holds

$$\|\nabla(u - u_h)\|_{L^2(\Omega)} \le \|\tau_h^{\mathrm{pw}}\|_{L^2(\Omega)} + \sqrt{\sum_{T \in \mathcal{T}} \frac{h_T^2}{\pi^2} \|f - \fint_T f \, dx\|^2_{L^2(T)}}.$$

Hint: You may use that the Poincaré constant on a convex domain ω can be bounded by $\mathrm{diam}(\omega)/\pi$.

PROBLEM 2.36. Consider the Dirichlet problem for the Laplacian with homogeneous boundary conditions on the unit square. with $f(x) = 2(x_1(1-x_1)+x_2(1-x_2))$ and exact solution $u(x) = x_1(x_1-1)x_2(x_2-1)$. Compute $\|\nabla u_h - \sigma_h\|_{L^2(\Omega)}$ on a sequence of mesh refinements and compare with the true errors.

PROBLEM 2.37. Implement the error estimator $\|\tau_h^{\mathrm{pw}}\|_{L^2(\Omega)}$ and test its performance for the setting of the previous Exercise.

PROBLEM 2.38. Prove the projection property $\nabla_h I_h = \Pi_0 \nabla$ for the nonconforming interpolation operator.

PROBLEM 2.39. Prove that $|||\cdot|||_h$ is a norm on $H_0^1(\Omega) + CR_0^1(\mathcal{T})$.

PROBLEM 2.40. Let T be a triangle and $v \in H^1(T)$ satisfy $\fint_E v\,ds = 0$ for an edge E of T. Prove

$$\|v\|_{L^2(T)} + h_T^{1/2}\|v\|_{L^2(E)} \le Ch_T\|\nabla v\|_{L^2(T)}$$

with a constant C that only depends on the shape regularity.

PROBLEM 2.41. Implement the Crouzeix–Raviart method for Ω and f as in Exercise 2.36 and produce experimental convergence history plots for the error in the $|||\cdot|||_h$ norm and the L^2 norm.

PROBLEM 2.42. Prove that the L^2 error of the Crouzeix–Raviart method is of order h^2 provided $u \in H_0^1(\Omega) \cap H^2(\Omega)$.

PROBLEM 2.43. Prove that the functions α_z, α_E for all interior vertices z and interior edges E form a basis of Z_h if Ω is simply connected. How needs the construction be modified for domains with holes?

PROBLEM 2.44. Implement the Crouzeix–Raviart method for the Stokes equations. As a test example, use the following data on the square $\Omega = (-1,1)^2$ (not the unit square): The right-hand side $f = 0$ is zero and the exact solution is

$$u(x_1,x_2) = \begin{pmatrix} 20x_1x_2^4 - 4x_1^5 \\ 20x_1^4x_2 - 4x_2^5. \end{pmatrix}$$

Choose the inhomogeneous Dirichlet data u_D according to u. Create convergence history plots for the $|||\cdot|||_h$ error in the u variable and the L^2 error in the p variable.

CHAPTER 3

Selected topics

3.1. Some details on Sobolev spaces and traces

We want to understand the origin of the notation $H^{1/2}(\partial\Omega)$ for the range of the trace operator and the connection to Sobolev scales.

3.1.1. Sobolev spaces of non-integer order. We begin by defining the spaces H^s for non-integer values of s.

DEFINITION 3.1 (Sobolev–Slobodeckij norm). Let $\Omega \subseteq \mathbb{R}^n$. For and $0 < s < 1$ and $v \in L^2(\Omega)$ we define

$$|v|_{H^s(\Omega)} := \left(\int_\Omega \int_\Omega \frac{|v(x) - v(y)|^2}{|x-y|^{n+2s}} \, dxdy \right)^{1/2} \qquad \in \mathbb{R} \cup \{\infty\}.$$

For a nonnegative integer $k \geq 0$ we define the space

$$H^{k+s}(\Omega) := \{v \in H^k(\Omega) : |\partial^\alpha v|_{H^s(\Omega)} < \infty \quad \text{for all multiindices with } |\alpha| = k\}$$

endowed with the *Sobolev–Slobodeckij norm*

$$\|v\|_{H^{k+s}(\Omega)} = \sqrt{\|v\|^2_{H^k(\Omega)} + \sum_{|\alpha|=k} |\partial^\alpha v|^2_{H^s(\Omega)}}.$$

♦

With this definition, we have a definition of the fractional-order space $H^{1/2}(\Omega)$. Since the boundary $\partial\Omega$ of our Lipschitz polytope Ω is a manifold, this does not directly give a definition of $H^{1/2}(\partial\Omega)$. In prior sections the latter space was already defined as the range of the trace operator, but for the moment we cancel that definition. The idea for defining $H^{1/2}(\partial\Omega)$ is to locally represent the boundary as the graph of a Lipschitz function, to flatten the boundary after localization with a suitable partition of unity, and to sum up the local $H^{1/2}$ norms of the transformed function. We recall the definition of a Lipschitz domain (first part of this lecture), where the open sets $U^1, \ldots, U^N$ cover a neighbourhood of $\partial\Omega$ and, after rotating and shifting the coordinate system, $U^j \cap \partial\Omega = \{(z, \gamma_j(z)) : z \in \tilde{U}^j\}$ is the graph of a Lipschitz function γ_j with the domain on one side of the graph. Here, $\tilde{U}^j \subseteq \mathbb{R}^{n-1}$ is the domain of γ_j. We also consider a corresponding functions $\eta_j \in C_c^\infty(U^j)$ that form a partition of unity on the boundary, $\sum_j \eta_j = 1$ on $\partial\Omega$.

DEFINITION 3.2 (H^s on the boundary). Let $\Omega \subseteq \mathbb{R}^n$ be an open bounded Lipschitz domain. We say that $u : \partial\Omega \to \mathbb{R}$ belongs to $H^s(\partial\Omega)$ if each function $u_j = (\eta_j u)(\cdot, \gamma_j(\cdot))$ belongs to $H^s(\tilde{U}^j)$. We define the (square of the) seminorm

$$|u|^2_{H^s(\partial\Omega)} := \sum_j \int_{\tilde{U}_j} \int_{\tilde{U}_j} \frac{|u_j(x) - u_j(y)|^2}{|x-y|^{n-1+2s}} dxdy.$$

(Note that x, y belong to $\mathbb{R}^{n-1}$). We define the norm $\|u\|_{H^s(\partial\Omega)} := (\|u\|^2_{L^2(\partial\Omega)} + |u|^2_{H^s(\partial\Omega)]})^{1/2}$. ♦

REMARK 3.3. The value of the norm (but not its finiteness) in the above definition depends on the choice of the U^j and η_j. ♦

What we shall prove in this section is that the space $H^{1/2}(\partial\Omega)$ equals the range of the trace operator, i.e., every $g \in H^{1/2}(\partial\Omega)$ is the trace of some $u \in H^1(\Omega)$ with $\|u\|_{H^1(\Omega)} \leq C\|g\|_{H^{1/2}(\partial\Omega)}$. Hence this alternative definition is equivalent to the one given above using the minimal extension norm.
We will not discuss traces of $H^s(\Omega)$ in detail, but what is important to observe is that functions from that space cannot have discontinuities on $(n-1)$-dimensional submanifolds if $s > 1/2$, but they can if $s < 1/2$, see Exercise 3.1. The case $s = 1/2$ is critical and it turns out that such functions can only have certain discontinuities.

EXAMPLE 3.4. A piecewise constant and discontinuous function u satisfies $u \notin H^{1/2}(\partial\Omega)$. But, for example, $u(t) = \log(|\log(|t|)|)$ belongs to

$$H^{1/2}(-1/\exp(1), 1/\exp(1)).$$

This will be proven later, cf. Exercise 3.6.

Generally for $v \in L^2(\Omega)$, we denote by $\tilde{v} \in L^2(\mathbb{R}^n)$ the extension by 0.

DEFINITION 3.5. Let $\Omega \subseteq \mathbb{R}^n$. We define

$$\widetilde{H}^{1/2}(\Omega) := \{v \in H^{1/2}(\Omega) : \tilde{v} \in H^{1/2}(\mathbb{R}^n)\}$$

with the norm

$$\|v\|_{\widetilde{H}^{1/2}(\Omega)} := \|\tilde{v}\|_{H^{1/2}(\mathbb{R}^n)}.$$

♦

We have $\|v\|_{H^{1/2}(\Omega)} \leq \|v\|_{\widetilde{H}^{1/2}(\Omega)}$ for any $v \in \widetilde{H}^{1/2}(\Omega)$, see Exercise 3.5.

DEFINITION 3.6. We denote by $H_0^s(\Omega)$ the closure of $C_c^\infty(\Omega)$ with respect to the H^s norm. We denote by $H^{-s}(\Omega)$ the dual of $H_0^s(\Omega)$. ♦

THEOREM 3.7 (density). *Let $\Omega \subseteq \mathbb{R}^n$ be an open and bounded Lipschitz domain and let $0 < s < 1$. Then, $H^s(\Omega)$ is a Banach space and we have $H^1(\Omega) \subseteq H^s(\Omega)$. The space $C^\infty(\overline{\Omega})$ is dense in $H^s(\Omega)$. We have*

$$H_0^s(\Omega) = \begin{cases} H^s(\Omega) & \text{if } 0 < s \leq 1/2 \\ \widetilde{H}^s(\Omega) & \text{if } 1/2 < s < 1. \end{cases}$$

PROOF. See for example [**Gri85**] or [**Dob10**]. □

REMARK 3.8. We stress the very important fact that $H^{1/2}(\Omega)$ is the closure of functions with compact support, but, at the same time, the elements in that space do not necessarily admit an $H^{1/2}$-regular extension by zero to the full space. ♦

3.1.2. The range of the trace operator.

LEMMA 3.9 (trace). *Let $\Omega \subseteq \mathbb{R}^n$ be an open and bounded Lipschitz domain. The trace operator is continuous as a map from $H^1(\Omega)$ to $H^{1/2}(\partial\Omega)$.*

PROOF. For simplicity we assume $n = 2$. Let $u \in C^1(\bar{\Omega})$. We localize the boundary with the sets U^j and the cutoff functions η_j and use Exercise 3.4. We can then assume without loss of generality that the support of u intersects the boundary such that $u|_{\partial\Omega}$ vanishes outside some $\Gamma \subseteq \partial\Omega$ which is the graph of a function γ_j over a subset of $\mathbb{R}^{n-1} = \mathbb{R}$, and $\Omega \subseteq \mathbb{R} \times \mathbb{R}_+$. We fix $x, y \in \mathbb{R}$ and define $\xi = (x-y)/2$ and $z = (\frac{1}{2}(x+y), |\xi|)$. We use the triangle inequality

$$|u(x,0) - u(y,0)| \leq |u(z) - u(x,0)| + |u(z) - u(y,0)|$$

and focus on the first term on the right-hand side. We use the fundamental theorem of calculus and obtain

$$|u(z) - u(x,0)| = |\int_0^1 \nabla u(x - t\xi, t|\xi|) \cdot \begin{pmatrix} -\xi \\ |\xi| \end{pmatrix} dt| \leq \sqrt{2}|\xi| \int_0^1 |\nabla u(x - t\xi, t|\xi|)|\, dt.$$

We square, divide by $|\xi|$ and integrate with respect to x and y. From symmetry in x and y we then obtain

$$|u(\cdot,0)|_{H^{1/2}(\Gamma)} \lesssim \int_0^1 \left(\int_\Gamma \int_\Gamma |\nabla u(x - t\xi, t|\xi|)|^2 \, dx\, dy \right)^{1/2} dt.$$

Here we have used Jensen's inequality $\int |f| \lesssim (\int f^2)^{1/2}$. Since u is compactly supported, we can replace the Γ in the integrals on the right-hand side by $\mathbb{R}$. We substitute with ξ

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} |\nabla u(x - t\xi, t|\xi|)|^2 \, dx\, dy &= 2 \int_{\mathbb{R}} \int_{\mathbb{R}} |\nabla u(x - t\xi, t|\xi|)|^2 \, dx\, d\xi \\ &= 2 \int_{\mathbb{R}} \int_{\mathbb{R}} |\nabla u(x, t|\xi|)|^2 \, dx\, d\xi. \end{aligned}$$

After changing coordinates $\xi \mapsto \xi/t$, we thus obtain

$$|u(\cdot,0)|_{H^{1/2}(\Gamma)} \lesssim \int_0^1 t^{-1/2} \|\nabla u\|_{L^2(\Omega)}\, dt \lesssim \|\nabla u\|_{L^2(\Omega)}.$$

This and density from Theorem 3.7 prove the continuity. □

Conversely, any $v \in H^{1/2}(\partial\Omega)$ admits a bounded extension to $\hat{v} \in H^1(\Omega)$.

LEMMA 3.10. *Let $\Omega \subseteq \mathbb{R}^n$ be an open and bounded Lipschitz domain. For every $v \in H^{1/2}(\partial\Omega)$ there exists an extension $\hat{v} \in H^1(\Omega)$ with $\|\hat{v}\|_{H^1(\Omega)} \leq C\|v\|_{H^{1/2}(\partial\Omega)}$ and $v = \hat{v}|_{\partial\Omega}$*

PROOF. Again, we will prove this in the simplified situation of two dimensions to keep the technicalities to a minimum. The principal mathematical argument, however, is the same in higher dimensions. As in the previous proof we may assume that $v \in \widetilde{H}^{1/2}(\Gamma)$ for a bounded interval $\Gamma \subseteq \mathbb{R}$, and in view of Exercise 3.2 we can assume that $\Gamma = \mathbb{R}$. We denote the coordinates of $\mathbb{R}^2$ by (x, y). We denote by ϕ the standard mollifier in $\mathbb{R}^2$ with support in the unit ball and unit integral, and set

$$\hat{v}(x,y) := \frac{1}{y} \int_{\mathbb{R}} \phi\left(\frac{z-x}{y}\right) v(z)\, dz, \qquad y > 0.$$

We compute (note that ϕ' has zero integral) the derivative and change coordinates,

$$\begin{aligned} \partial_x \hat{v}(x,y) &= -\frac{1}{y^2} \int_{\mathbb{R}} \phi'\left(\frac{z-x}{y}\right) (v(z) - v(x))\, dz \\ &= \frac{1}{y} \int_{|z|<1} \phi'(z)(v(x) - v(x+yz))\, dz. \end{aligned}$$

After squaring and integrating and observing that ϕ' is bounded, we compute

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}_+} |\partial_x \hat{v}(x,y)|^2 \, dx\, dy &\lesssim \int_{\mathbb{R}} \int_{\mathbb{R}_+} y^{-2} \int_{|z|<1} |v(x) - v(x+yz)|^2 \, dz\, dx\, dy \\ &\lesssim \int_{\mathbb{R}} \int_{\mathbb{R}_+} y^{-3} \int_{|x-w|<y} |v(x) - v(w)|^2 \, dw\, dx\, dy \\ &\lesssim \int_{\mathbb{R}} \int_{\mathbb{R}} |v(x) - v(w)|^2 \int_{|x-w|}^{\infty} y^{-3} \, dy\, dx\, dw. \end{aligned}$$

The y-integral equals $2^{-1}|x-w|^{-2}$, and therefore we have shown $\|\partial_x \hat{v}\|_{L^2(\Omega)} \lesssim |v|_{H^{1/2}(\Gamma)}$.

We next bound the derivative $\partial_y \hat{v}$. We observe that integration by parts implies

$$\int_{\mathbb{R}} \phi\left(\frac{z-x}{y}\right) dz + \int_{\mathbb{R}} \phi'\left(\frac{z-x}{y}\right) \frac{z-x}{y} dz = 0.$$

We can therefore compute

$$\begin{aligned}\partial_y \hat{v}(x,y) = &- y^{-2} \int_{\mathbb{R}} \phi\left(\frac{z-x}{y}\right)(v(z)-v(x))dz \\ &- y^{-2} \int_{\mathbb{R}} \phi'\left(\frac{z-x}{y}\right) \frac{z-x}{y}(v(z)-v(x))dz.\end{aligned}$$

The integrals are bounded in a similar fashion as before. □

The preceding two results show that both definitions of $H^{1/2}(\partial\Omega)$ given in these notes are equivalent, and so are their norms.

LEMMA 3.11. *Let $\Omega \subseteq \mathbb{R}^n$ be an open bounded Lipschitz domain. The derivative ∂_{x_j} continuously maps $H^{1/2}(\Omega)$ to the dual space $[\widetilde{H}^{1/2}(\Omega)]^*$.*

PROOF. For the ease of notation we consider $n = 1$ and denote with x, y the Cartesian coordinates of $\mathbb{R}^2$. We consider functions $v \in H^{1/2}(\Omega)$ (with some continuation to $H^{1/2}(\mathbb{R})$). and $w \in \widetilde{H}^{1/2}(\Omega)$, which admits a bounded extension by zero to an object of $H^{1/2}(\mathbb{R})$. From previous proofs we know that these functions have bounded extensions $\hat{v} \in H^1(\mathbb{R} \times \mathbb{R}_+)$ and $\hat{w} \in H^1(\mathbb{R} \times \mathbb{R}_+)$. We obtain from integration by parts that

$$\int_\Omega \partial_x \hat{v}(x,y)\hat{w}(x,y)\,dx = -\int_y^\infty \int_{\mathbb{R}^n} (\partial_x \hat{v}(x,s)\partial_y \hat{w}(x,s) - \partial_y \hat{v}(x,s)\partial_x \hat{w}(x,s))\,ds\,dx.$$

For $y \to 0$ we obtain that

$$\int_\Omega \partial_x v(x) w(x)\,dx \lesssim \|\nabla \hat{v}\|_{H^1(\mathbb{R}\times\mathbb{R}_+)} \|\nabla \hat{w}\|_{H^1(\mathbb{R}\times\mathbb{R}_+)} \lesssim \|v\|_{H^{1/2}(\Omega)} \|w\|_{\widetilde{H}^{1/2}(\Omega)}.$$

for any such pair of functions. □

The previous result is sharp in the sense that the partial derivative does not map $H^{1/2}(\Omega)$ to $H^{-1/2}(\Omega)$ for a bounded Lipschitz domain Ω, see Exercise 3.6. But we have that the tangential derivative maps $H^{1/2}(\partial\Omega)$ to $H^{-1/2}(\partial\Omega)$, which was already used in Example 2.30, see also Exercise 3.7.

3.2. Corner singularities in planar domains

3.2.1. Setting. This section provides a brief introduction to the regularity theory of elliptic second-order boundary value problems in Lipschitz polygons ("polygons" for short), that is, in open and bounded domains whose boundary can locally be represented as the graph of a piecewise affine function. As a simplification of the general situation presented in [**Gri92**, Chapter 2], we consider the Dirichlet Laplacian as a model case,

$$-\Delta u = f \quad \text{in } \Omega \qquad \text{and} \qquad u = 0 \text{ on } \partial\Omega.$$

In a Hilbert space setting, this problem has a unique solution Sobolev space $H^1_0(\Omega)$ for any $f \in H^{-1}(\Omega)$. More precisely, the operator

$$-\Delta : H^1_0(\Omega) \to H^{-1}(\Omega)$$

is an isomorphism. If we restrict our attention to right-hand sides f from $L^2(\Omega)$, the range of the solution operator $(-\Delta)^{-1}|_{L^2(\Omega)}$ is a subspace of $H^1_0(\Omega)$, and regularity theory tries to find characterizations of this subspace. It is known that such solutions enjoy H^2 regularity in the interior that can be extended up to the boundary provided the latter is sufficiently smooth, say it belongs to the class C^2.

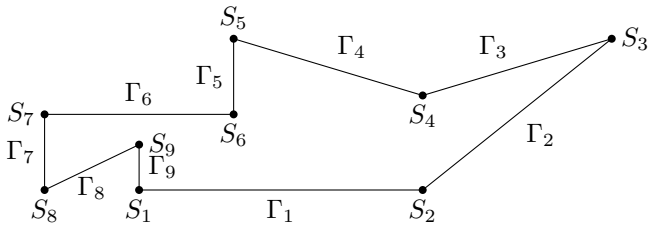


FIGURE 1. Our notation for a polygon.

In this case, it can be shown by local flattening and reflection techniques [**Eva10**] that $u \in H_0^1(\Omega) \cap H^2(\Omega)$ whenever $f \in L^2(\Omega)$. In domains with corners (such as polygons) this result is not generally true.

EXAMPLE 3.12. Let $\Omega := \{(r, \theta) : 0 < r < 1 \text{ and } 0 < \theta < 3\pi/2\}$ denote the sector domain (r and θ are the usual polar coordinates). Then

$$u(r, \theta) = r^{2/3} \sin(2\theta/3)$$

belongs to $H^1(\Omega)$, satisfies zero boundary conditions near $(0, 0)$, but does not belong to $H^2(\omega)$ for any open subdomain $\omega \subset \Omega$ such that $(0, 0) \in \bar{\omega}$. On the other hand we have that $\Delta u = 0$, which belongs to $L^2(\Omega)$.

It will turn out that functions as in this example will describe characteristic singularities near corners. We shall prove that the operator $-\Delta$ maps $H_0^1(\Omega) \cap H^2(\Omega)$ to a closed subspace of $L^2(\Omega)$, whose orthogonal complement N has a dimension related to the corners of the domain. If Ω has finitely many corners, then N is finite-dimensional. This is the main decomposition theorem. Moreover, this characterization makes it possible to precisely predict the regularity of the solution using fractional Sobolev spaces. In the above example, the solution satisfies the regularity

$$u \in H^{5/3-\delta}(\Omega) \quad \text{for any } \delta > 0.$$

We will study how regularity in the fractional-order Sobolev spaces $H^s(\Omega)$ for $0 < s < 1$ is related to the corners of the domain.
In what follows, $\Omega \subset \mathbb{R}^2$ is a bounded and open polygon with (for simplicity) finitely many corners. Thus, there exists a positive integer M such that the boundary consists of M many straight line segments $(\Gamma_j : j = 1, \ldots, M)$ meeting at corners $(S_j : 1 = 1, \ldots, M)$ where $S_j := \Gamma_j \cap \Gamma_{j+1}$, see Figure 1 for an illustration.
We consider the space

$$H(\Delta, \Omega) := \{v \in L^2(\Omega) : \Delta v \in L^2(\Omega)\}$$

with the norm

$$\|v\|_{H(\Delta,\Omega)} = \left(\|v\|^2_{L^2(\Omega)} + \|\Delta v\|^2_{L^2(\Omega)}\right)^{1/2}.$$

Note that first-order partial derivatives of functions from this space will in general only exist as distributions, but not as L^2 functions. Nevertheless, we can give a meaning to traces of functions from $H(\Delta, \Omega)$. We note that the outward pointing unit normal vector ν to $\partial\Omega$ exists almost everywhere on $\partial\Omega$ (namely in the interior of any of the segments Γ_j).

LEMMA 3.13. *Consider the space*

$$W := H^2(\Omega) \cap H_0^1(\Omega) \subset H^2(\Omega).$$

The trace mapping $\gamma : H^2(\Omega) \to W^*$ *defined by*

$$v \mapsto \left[w \mapsto \int_{\partial\Omega} v \frac{\partial w}{\partial \nu}\, ds \right] =: \langle \gamma v, \cdot \rangle \in W^*$$

has a unique continuous extension to a linear map from $H(\Delta, \Omega)$ *to* W^*.

PROOF. Let $v \in H^2(\Omega)$ and $w \in W$. Integration by parts (applied twice) shows

$$\int_\Omega v \Delta w \, dx = \int_{\partial\Omega} v \frac{\partial w}{\partial \nu}\, ds - \int_{\partial\Omega} w \frac{\partial v}{\partial \nu}\, ds + \int_\Omega w \Delta v \, dx.$$

Since w vanishes on the boundary, the second integral on the right-hand side equals zero. This and the Cauchy inequality establish

$$\begin{aligned} \int_{\partial\Omega} v \frac{\partial w}{\partial \nu}\, ds &= \int_\Omega v \Delta w \, dx - \int_\Omega w \Delta v \, dx \\ &\le \|v\|_{L^2(\Omega)} \|\Delta w\|_{L^2(\Omega)} + \|w\|_{L^2(\Omega)} \|\Delta v\|_{L^2(\Omega)} \\ &\le C \|v\|_{H(\Delta,\Omega)} \|w\|_{H^2(\Omega)}. \end{aligned}$$

The result follows from density of $H^2(\Omega)$ in $H(\Delta, \Omega)$ (see Exercise 3.12). □

REMARK 3.14. We interpret γu as a boundary trace for $u \in H(\Delta, \Omega)$ and write $u|_{\partial\Omega}$ instead of γu. ♦

3.2.2. The decomposition theorems. Recall the notation $W := H^2(\Omega) \cap H_0^1(\Omega)$. We consider the Laplacian as an operator $\Delta : W \to L^2(\Omega)$. Injectivity and closed range property of Δ follow from Exercise 3.9. We are interested in

$$N := \{ v \in L^2(\Omega) : \forall w \in W \ (\Delta w, v)_{L^2(\Omega)} = 0 \} = (\Delta W)^\perp.$$

These are the right-hand sides leading to singular solutions to the Laplacian.

LEMMA 3.15. *We have* $v \in N$ *if and only if* $v \in H(\Delta, \Omega)$ *and*

$$\Delta v = 0 \text{ in } \Omega \qquad \text{and} \qquad v|_{\partial\Omega} = 0 \text{ in the sense of traces of } H(\Delta, \Omega).$$

PROOF. The proof is left to the reader as an exercise. □

LEMMA 3.16. *Let* $v \in N$ *and let* $U \subset \bar{\Omega}$ *denote any neighbourhood of the corners* $\{S_j\}$. *Then* $v \in C^\infty(\bar{\Omega} \setminus U)$.

PROOF. This is the classical interior regularity result, see [**Eva10**]. □

Consider the corner number j with angle ω_j and the operator

$$\Lambda_j : H^2(0, \omega_j) \cap H_0^1(0, \omega_j) \to L^2(0, \omega_j)$$

defined by

$$\Lambda_j \varphi = -\varphi''.$$

We know from the spectral theory of self-adjoint compact operators that Λ_j has a discrete spectrum with nonnegative eigenvalues $\lambda_{j,m}^2$ $(m = 1, 2, 3, \dots)$. The corresponding L^2-normalized eigenfunctions are denoted by $\varphi_{j,m}$. It is well known that

$$\lambda_{j,m} = m\pi/\omega_j \quad \text{and} \quad \varphi_{j,m}(\theta) = \sqrt{2/\omega_j} \sin(\theta \lambda_{j,m}).$$

Given any corner S_j we denote the polar coordinates with origin S_j by (r_j, θ_j). We choose $\rho_j > 0$ small enough such that $D_{\rho_j} := \Omega \cap \{0 < r_j < \rho_j\}$ does not intersect with parts of $\partial\Omega$ other than $\Gamma_j \cup \Gamma_{j+1}$. We will sometimes use cut-off functions $\eta_j \in C^\infty(\bar{\Omega})$, $j = 1, \dots, M$ with mutually disjoint supports and the property

$$\eta_j = \begin{cases} 1 & \text{in an open neighbourhood of } S_j \\ 0 & \text{outside } D_{\rho_j}. \end{cases}$$

We now fix one corner $S_j \equiv S$ and denote the polar coordinates with origin S by (r, θ). We write $\rho = \rho_j$ as well as $\lambda_m := \lambda_{j,m}$ and $\varphi_m := \varphi_{j,m}$, $\omega := \omega_j$.
The representation of the Laplacian in polar coordinates shows that any $v \in N$ satisfies

$$\frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} + \frac{1}{r^2}\frac{\partial^2 v}{\partial \theta^2} = 0 \quad \text{for } 0 < \theta < \omega, 0 < r < \rho.$$

It can be shown that v has zero boundary conditions away from S_j (prove this as an exercise), see Lemma 3.16. For any $0 < r < \rho$, we have

$$v(r, \theta) \in H^2(0, \omega_j) \quad \text{(as a function of } \theta)$$

and thus

$$\frac{\partial^2 v}{\partial r^2} + \frac{1}{r}\frac{\partial v}{\partial r} - \frac{1}{r^2}\Lambda_j v = 0 \quad \text{for } 0 < r < \rho. \tag{28}$$

LEMMA 3.17. *Let $v \in C^\infty((0,\rho); H^2(0,\omega_j) \cap H_0^1(0,\omega_j))$ solve (28) and assume $v \in L^2(D_\rho)$. Then there exist real numbers α_m, β_m with*

$$|\alpha_m| \le L m^{1/2} \rho^{-(\lambda_m+1)}$$

(L only dependent on v) such that

$$v(r,\theta) = \sum_{m\ge 1} \alpha_m r^{\lambda_m}\varphi_m(\theta) + \sum_{0<\lambda_m<1} \beta_m r^{-\lambda_m}\varphi_m(\theta).$$

PROOF. The functions φ_m form a complete orthonormal system of $L^2(0,\omega)$. Thus

$$v(r,\theta) = \sum_{m\ge 1} v_m(r)\varphi_m(\theta) \quad \text{with the coefficient} \quad v_m(r) = \int_0^\omega v(r,\theta)\varphi_m(\theta)d\theta.$$

The differential equation implies

$$v_m''(r) + r^{-1}v_m'(r) - \lambda_m^2 r^{-2} v_m(r) = 0 \quad \text{for } 0 < r < \rho.$$

This ODE has the following solutions (Exercise 3.10)

$$v_m(r) = \alpha_m r^{\lambda_m} + \beta_m r^{-\lambda_m} \quad \text{for } \lambda_m > 0 \quad \text{(relevant here)}$$
$$v_m(r) = \alpha_m + \beta_m \log(r) \quad \text{for } \lambda_m = 0 \quad \text{(not relevant here).}$$

Squaring the coefficient relation, integrating, and using Cauchy's inequality implies

$$\int_0^\rho |v_m(r)|^2 r dr = \int_0^\rho |\int_0^\omega v(r,\theta)\varphi_m(\theta)d\theta|^2 r dr \le \int_0^\rho\int_0^\omega |v(r,\theta)|^2 d\theta r dr$$
$$= \|v\|^2_{L^2(D_\rho)} < \infty.$$

Thus in case $\lambda_m \ge 1$, we see that $\beta_m = 0$. Furthermore, if $\lambda_m \ge 1$, we see that

$$\frac{|\alpha_m|^2}{2\lambda_m + 2}\rho^{2\lambda_m+2} = |\alpha_m|^2 \int_0^\rho r^{2\lambda_m+1} dr \le \|v\|^2_{L^2(D_\rho)}.$$

□

THEOREM 3.18. *The dimension of N equals*

$$\sum_j \mathrm{card}\{\lambda_{j,m} : 0 < \lambda_{j,m} < 1\}.$$

PROOF. STEP 1. We begin by considering a fixed corner (number j) and the related eigenvalues λ_m and eigenfunctions φ_m. Let m be such that $\lambda_m \in (0,1)$. Recall the localization function $\eta \equiv \eta_j$ and the polar coordinates (r,θ) related to this corner. We define the function

$$u_m := \eta r^{-\lambda_m}\varphi_m(\theta).$$

We obviously have that $u_m \in H(\Delta, \Omega)$ (prove this as an exercise) with (generalized) zero boundary conditions.. We can thus solve for $v_m \in H_0^1(\Omega)$ with $\Delta v_m = \Delta u_m$ and set $\sigma_m := u_m - v_m$. We then have by construction that $\sigma_m \in H(\Delta, \Omega)$ and $\sigma_m|_{\partial\Omega} = 0$, furthermore $\Delta\sigma_m = 0$. By Lemma 3.16 we thus have $\sigma_m \in N$. Therefore we have shown that there exists $\sigma_m \in N$ such that

$$\sigma_m - \eta r^{-\lambda_m}\varphi_m(\theta) \in H^1(\Omega).$$

STEP 2. Let $v \in N$. We have seen in the lemma that near our corner (number j) we have

$$v(r,\theta) - \sum_{m\geq 1} \alpha_m r^{\lambda_m}\varphi_m(\theta) - \sum_{0<\lambda_m<1} \beta_m r^{-\lambda_m}\varphi_m(\theta) = 0.$$

We have seen that $r^{-\lambda_m}\varphi_m(\theta)$ and σ_m only differ by an $H^1(\Omega)$ function, thus upon substituting we obtain

$$v(r,\theta) - \sum_{m\geq 1} \alpha_m r^{\lambda_m}\varphi_m(\theta) - \sum_{0<\lambda_m<1} \beta_m \sigma_m \in H^1(D_\rho).$$

It is proved as an exercise that (with the help of the bounds on α_m from Lemma 3.17)

$$\sum_{m\geq 1} \alpha_m r^{\lambda_m}\varphi_m(\theta) \in H^1(D_{\rho'}) \qquad \text{for any } 0 < \rho' < \rho.$$

Consequently, we infer that

$$v(r,\theta) - \sum_{0<\lambda_m<1} \beta_m \sigma_m \in H^1(D_{\rho'}).$$

STEP 3. The interior regularity from Lemma 3.16 then shows that, in global notation, we have

$$w := v - \sum_j \sum_{0<\lambda_{j,m}<1} \beta_{j,m}\sigma_{j,m} \in H^1(\Omega).$$

On the other hand, since $w \in N \cap H^1(\Omega)$, we know by Lemma 3.16 that $w \in H^1(\Omega)$ is harmonic with zero boundary conditions. Thus, $w = 0$ and

$$v = \sum_j \sum_{0<\lambda_{j,m}<1} \beta_{j,m}\sigma_{j,m}.$$

□

For any corner S_j of the domain Ω we define the "singularity function" τ_j by

$$\tau_j(r_j,\theta_j) = \eta_j(r_j) r_j^{\lambda_{j,1}}\varphi_{j,1}(\theta_j).$$

These functions have the following properties.

LEMMA 3.19. *The functions τ_j $(j = 1,\dots,M)$ satisfy*

$$\tau_j \in H(\Delta,\Omega) \quad \textit{and} \quad \tau_j|_{\partial\Omega} = 0.$$

The functions $(\Delta\tau_j : j = 1,\dots,M)$ are linearly independent. If $\lambda_{j,1} < 1$, then $\Delta\tau_j$ is not orthogonal to the space N.

PROOF. Exercise 3.14. □

THEOREM 3.20. *Let $\Omega \subset \mathbb{R}^2$ be a connected and open polygonal domain and $f \in L^2(\Omega)$, and denote by $u \in H_0^1(\Omega)$ the solution to the Poisson equation*

$$-\Delta u = f \quad \textit{in } \Omega \quad \textit{and} \quad u = 0 \quad \textit{on } \partial\Omega.$$

Then there exist real coefficients $(c_1,\dots,c_M) \in \mathbb{R}^M$ such that

$$u - \sum_{\substack{j \text{ with} \\ \omega_j > \pi}} c_j\tau_j \in H^2(\Omega).$$

PROOF. We know that $\lambda_{j,m} = m\pi/\omega_j$. Thus

$$\lambda_{j,m} < 1 \quad \text{if and only if} \quad \omega_j > \pi \text{ and } m = 1.$$

Theorem 3.18 thus teaches us that

$$\dim N = \{j : \omega_j > \pi\}.$$

The functions τ_j for $\omega_j > \pi$ are linearly independent and thus, by a dimension argument, form a basis of N. Consequently, the space $L^2(\Omega)$ is spanned by the range of $\Delta((H_0^1(\Omega) \cap H^2(\Omega))$ and the functions $\Delta\tau_j$. Thus, given $f \in L^2(\Omega)$, there exists $w \in H_0^1(\Omega) \cap H^2(\Omega)$ and coefficients c_j such that

$$f = \Delta w + \sum_{\substack{j \text{ with} \\ \omega_j > \pi}} c_j \Delta\tau_j.$$

The assertion of the theorem follows from the uniqueness of the solution to the variational problem (i.e., apply Δ^{-1} on both sides). □

We end this section with a quantification of regularity in Sobolev spaces of fractional order.

THEOREM 3.21. *Let $\Omega \subset \mathbb{R}^2$ be a connected and open polygonal domain and $f \in L^2(\Omega)$. The solution $u \in H_0^1(\Omega)$ to the Poisson equation*

$$-\Delta u = f \quad \text{in } \Omega \quad \text{and} \quad u = 0 \quad \text{on } \partial\Omega$$

satisfies

$$u \in H^{1+s}(\Omega) \quad \text{for any } s < \min\{1, \min_{j=1,\ldots,M} \frac{\pi}{\omega_j}\}.$$

PROOF. Details are worked out in Exercise 3.13. □

3.A. Problems

PROBLEM 3.1. Let $u(x) = \operatorname{sign}(x_1)$ and let $\Omega = (-1,1)^n$ denote the hypercube in $\mathbb{R}^n$. Prove that $u \in H^s(\Omega)$ if $0 < s < 1/2$ and that $u \notin H^s(\Omega)$ if $1/2 \le s < 1$.

PROBLEM 3.2. Let $0 < s < 1$ and let $u \in H^s(\Omega)$ have compact support in Ω. Denote $\delta = \operatorname{dist}(\operatorname{supp}(u), \partial\Omega)$ and let $\tilde{u}$ denote the continuation of u by zero to $\mathbb{R}^n$. Prove that $\tilde{u} \in H^s(\mathbb{R}^n)$ and

$$\|\tilde{u}\|^2_{H^s(\mathbb{R}^n)} \le C(1 + s^{-1}\delta^{-2s})\|u\|^2_{H^s(\Omega)}.$$

Hint: Lemma 6.34 in [**Dob10**].

PROBLEM 3.3. Let $u : \mathbb{R}^n \to \mathbb{R}$ be of class C^1 and let $x, y \in \mathbb{R}^n$. Prove that

$$u(z) - u(x) = \int_0^1 \nabla(tz + (1-t)x) \cdot (z - x)\, dt.$$

PROBLEM 3.4. Convince yourself that locally flattening the boundary of a Lipschitz domain preserves the H^1 property. Consult Lemma 6.6 from [**Dob10**]

PROBLEM 3.5. Let $u \in \widetilde{H}^{1/2}(\Omega)$. Prove that

$$\|u\|^2_{\widetilde{H}^{1/2}(\Omega)} = \|u\|^2_{H^{1/2}(\Omega)} + 2\int_\Omega |u(x)|^2 \int_{\mathbb{R}^n \setminus \Omega} |x - y|^{-n-1}\, dy\, dx.$$

PROBLEM 3.6. Prove that the bilinear form $(v, w) \mapsto \int_0^1 v'(x)w(x)\, dx$ does not possess a continuous extension to $H^{1/2}((0,1)) \times H^{1/2}((0,1))$. *Hint: Consider the function* $\log(|\log(x/\exp(1))|)$.

PROBLEM 3.7. Let $\Omega \subseteq \mathbb{R}^2$ be a bounded open Lipschitz polygon. Prove that the tangential derivative ∂_s is a continuous map from $H^{1/2}(\partial\Omega)$ to $H^{-1/2}(\partial\Omega)$.

PROBLEM 3.8. On the L-shaped domain $\Omega = (-1,1)^2 \setminus ([0,1]\times[-1,0])$ we are given the Dirichlet boundary $\Gamma_D = \{0\}\times[-1,0]\cup[0,1]\times\{0\}$ and the Neumann boundary $\Gamma_N = \partial\Omega \setminus \Gamma_D$. For boundary data $u_D = 1$ on Γ_D, and Neumann data

$$g(x,y) = \frac{2}{3} r^{-1/3} \times \begin{cases} \cos(\phi)\sin(2\phi/3) - \sin(\phi)\cos(2\phi/3) & \text{if } x = 1 \\ \sin(\phi)\sin(2\phi/3) + \cos(\phi)\cos(2\phi/3) & \text{if } y = 1 \\ -\cos(\phi)\sin(2\phi/3) + \sin(\phi)\cos(2\phi/3) & \text{if } x = -1 \\ -\sin(\phi)\sin(2\phi/3) - \cos(\phi)\cos(2\phi/3) & \text{if } y = -1 \end{cases}$$

in polar coordinates (r,ϕ), and $f = 0$, solve the mixed boundary value problem for the Laplacian with the mixed Raviart–Thomas FEM. The exact solution is given by $u(r,\phi) = 1 + r^{2/3}\sin(2\phi/3)$. Plot the convergence history for the L^2 norm of $u - u_h$ as well as $\sigma - \sigma_h$ and on $\Pi_h u - u_h$.

PROBLEM 3.9. Let $\Omega \subset \mathbb{R}^2$ be a connected and open polygonal domain. Prove that every $u \in H^2(\Omega) \cap H_0^1(\Omega)$ satisfies the identity

$$\|\Delta u\|^2 = \|D^2 u\|^2.$$

Conclude that $\Delta : H^2(\Omega) \cap H_0^1(\Omega) \to L^2(\Omega)$ is injective with closed range. Furthermore, prove that exists a constant $C(\Omega)$ such that every $u \in H^2(\Omega) \cap H_0^1(\Omega)$ satisfies

$$\|u\|_{H^2(\Omega)} \le C(\Omega)\|\Delta u\|.$$

PROBLEM 3.10. Consider the ODE

$$v''(r) + r^{-1}v'(r) - \lambda^2 r^{-2} v(r) = 0 \qquad 0 < r < \rho$$

for some nonnegative real number λ. Prove that the solution is given by

$$v(r) = \begin{cases} \alpha r^\lambda + \beta r^{-\lambda} & \text{if } \lambda > 0 \\ \alpha + \beta\log(r) & \text{if } \lambda = 0 \end{cases}$$

with real numbers α, β. *Hint: The ODE is called Cauchy–Euler equation.*

PROBLEM 3.11. We consider $v \in H(\Delta,\Omega)$ for a polygon Ω. Let $\Gamma \subseteq \partial\Omega$ be a straight segment of the boundary. Prove that $v|_\Gamma \in [\widetilde{H}^{1/2}(\Gamma)]^*$. *Hint: You may use* [**Gri85**] *that the trace of $\partial\cdot/\partial\nu$ is continuous and onto from $H_0^1(\Omega)\cap H^2(\Omega)$ to $\widetilde{H}^{1/2}(\Omega')$ for any convex polygon Ω'.*

PROBLEM 3.12. Let $\Omega \subset \mathbb{R}^2$ be a connected and open polygonal domain. Prove that $H^2(\Omega)$ is dense in $H(\Delta,\Omega)$, but $H_0^1(\Omega)\cap H^2(\Omega)$ is not dense in $H_0^1(\Omega)\cap H(\Delta,\Omega)$.

PROBLEM 3.13. Show that in two dimensions and for $0 < s, \alpha < 1$, we have $r^\alpha \in H^{1+s}(\Omega)$ if and only if $s < \alpha$. Prove Theorem 3.21.

PROBLEM 3.14. Prove Lemma 3.19.

Bibliography

[Bar16] Sören Bartels. *Numerical approximation of partial differential equations.*, volume 64 of *Texts Appl. Math.* Springer, Cham, 2016. Online im Netz der Uni Jena verfügbar.

[BBF13] Daniele Boffi, Franco Brezzi, and Michel Fortin. *Mixed Finite Element Methods and Applications*, volume 44 of *Springer Series in Computational Mathematics*. Springer, Heidelberg, 2013.

[Bra07] D. Braess. *Finite Elements. Theory, Fast Solvers, and Applications in Elasticity Theory.* Cambridge University Press, Cambridge, third edition, 2007.

[BS08] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.

[Cia78] Philippe G. Ciarlet. *The Finite Element Method for Elliptic Problems*, volume 4 of *Studies in Mathematics and its Applications*. North-Holland, Amsterdam, 1978.

[Dob10] Manfred Dobrowolski. *Angewandte Funktionalanalysis. Funktionalanalysis, Sobolev-Räume und elliptische Differentialgleichungen*. Berlin: Springer, 2nd revised and extended ed. edition, 2010.

[Eva10] Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2. edition, 2010.

[Gri85] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*, volume 24 of *Monographs and Studies in Mathematics*. Pitman, Boston, MA, 1985.

[Gri92] P. Grisvard. *Singularities in Boundary Value Problems*, volume 22 of *Recherches en Mathématiques Appliquées*. Masson, Paris, 1992.