Appendix A
Problems and Projects

A.1 Finite Difference Method

A.1.1 Transport Equation

Exercise A.1.1 Derive a partial differential equation that describes the transport of a substance through a long, thin tube that allows for the injection of a substance at any time \( t \in [0, T] \) and any position \( x \in \mathbb{R} \) described through a function \( f(t, x) \) that specifies the number of injected particles per unit volume.

Exercise A.1.2

(i) Let \( f \in C(\mathbb{R}) \) and such that
\[
\int_{x_1}^{x_2} f(x) \, dx = 0
\]
for all \( x_1 \leq x_2 \). Show that \( f = 0 \).

(ii) Show that in the derivation of the transport equation, we have that
\[
\int_{x_1}^{x_2} u(t, x) \, dx = \int_{x_1}^{x_2} u(t + \tau, x + a\tau) \, dx,
\]
and conclude that \( u(t, x) = u(t + \tau, x + a\tau) \) for all \( t \in [0, T] \), \( x \in \mathbb{R} \), and \( \tau > 0 \).
Exercise A.1.3

(i) Prove the following estimates for difference quotients:

\[ |\partial^\pm_\nu u(x_j) - u'(x_j)| \leq \frac{\Delta \nu}{2} \|u''\|_{C([0,1])}, \]

\[ |\partial_\nu u(x_j) - u'(x_j)| \leq \frac{\Delta \nu^2}{6} \|u''\|_{C([0,1])}, \]

\[ |\partial^+ \partial^- u(x_j) - u''(x_j)| \leq \frac{\Delta \nu^2}{12} \|u''\|_{C([0,1])}. \]

Show that these estimates do not hold if \( u \) does not satisfy the required differentiability properties.

(ii) Show that \( \partial^+ \partial^- = \partial^- \partial^+ \).

(iii) Prove an error estimate for the difference \( \partial^+ \partial^+ u(x_j) - u''(x_j) \).

Exercise A.1.4 Let \( a < 0 \) and consider the numerical scheme \( \partial_t U^k_j + a \partial^+_x U^k_j = 0 \). Show that the scheme is stable under appropriate conditions on \( \Delta t \) and \( \Delta x \) and prove an error estimate.

Exercise A.1.5 Let \( u_0 \in C^2([0,1]) \) and let \( \tilde{u}_0 \) denote its trivial extension by zero to \( \mathbb{R} \).

(i) Find conditions on \( u_0 \) that guarantee \( \tilde{u}_0 \in C^2(\mathbb{R}) \).

(ii) Show that the solution of the transport problem \( \partial_t u + a \partial_x u = 0 \) with \( u(t,0) = 0 \) and \( u(0,x) = u_0(x) \) satisfies \( u \in C^2([0,T] \times [0,1]) \), and that

\[ \|\partial^2_t u(t,\cdot)\|_{C([0,1])} = a^{-2} \|\partial^2_t u(t,\cdot)\|_{C([0,1])} = \|u''\|_{C([0,1])} \]

for all \( t \in [0,T] \).

Exercise A.1.6 Show that the upwinding scheme for the transport equation is equivalent to the scheme

\[ \partial^+_t U^k_j + a_j^k \partial_x U^k_j = |a_j^k| \Delta x \partial^+_x \partial^-_x U^k_j, \]

and discuss the incorporation of boundary conditions.

Exercise A.1.7

(i) Show by constructing appropriate initial data that the difference scheme \( U^{k+1}_j = U^k_j + \mu(U^k_j - U^{k-1}_j) \) with \( \mu = a \Delta t/\Delta x \) is unstable if \( \mu > 1 \).

(ii) Check the CFL condition and the estimate \( \sup_{j=0,...,J} |U^{k+1}_j| \leq \sup_{j=0,...,J} |U^k_j| \) of the following difference schemes for the transport equation:

\[ \partial^+_t U^k_j - \partial^-_x U^k_j = 0, \quad \partial^+_t U^k_j + \partial^+_x U^k_j = 0, \quad \partial^+_t U^k_j + \partial^-_x U^k_j = 0. \]
Exercise A.1.8

(i) Show that the functions \( \phi_\ell(x) = e^{ikx}, x \in [-\pi, \pi], \ell \in \mathbb{Z}, \) define an orthonormal system in \( L^2(-\pi, \pi), \) i.e., for all \( \ell, m \in \mathbb{Z}, \) we have

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} \phi_\ell(x)\overline{\phi_m(x)} \, dx = \delta_{\ell m}.
\]

(ii) For \( f \in L^2(-\pi, \pi) \) and \( \ell \in \mathbb{Z} \) set

\[
f_\ell = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)\overline{\phi_\ell(x)} \, dx.
\]

Prove that

\[
\frac{1}{2\pi} \int_{-\pi}^{\pi} |f|^2 \, dx = \sum_{\ell \in \mathbb{Z}} |f_\ell|^2.
\]

Exercise A.1.9 Let \( g \in C([-\pi, \pi]) \) be such that

\[
\int_{-\pi}^{\pi} |fg| \, dx \leq \int_{-\pi}^{\pi} |f| \, dx
\]

for all \( f \in L^1(-\pi, \pi). \) Show that \( |g(x)| \leq 1 \) for all \( x \in [-\pi, \pi]. \) Is it sufficient to assume that \( g \in L^2(-\pi, \pi)? \)

Exercise A.1.10 Let \( u \) solve the partial differential equation \( \partial_t u + a(t, x)\partial_x u = 0. \)

(i) Show that \( u \) is constant along curves \( (t, y(t)) \) for solutions of the initial boundary value problems \( y'(t) = a(t, y(t)), y(0) = x_0, \) called characteristics.

(ii) Determine the characteristics for the equation \( \partial_t u + tx\partial_x u = 0, \) i.e., for \( a(t, x) = tx, \) sketch them, and determine the solution for the initial condition \( u_0(x) = \sin(x). \)

Quiz A.1.1 Decide for each of the following statements whether it is true or false.
You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True or False</th>
</tr>
</thead>
<tbody>
<tr>
<td>The transport equation describes the motion of a substance in a motionless fluid</td>
<td>True</td>
</tr>
<tr>
<td>The total amount of substance in the transport problem is conserved</td>
<td>True</td>
</tr>
<tr>
<td>For a ( C^2 ) function ( u, ) the central difference quotient ( \widehat{\partial} ) provides a more accurate approximation of the derivative than the one-sided difference quotients ( \partial^\pm )</td>
<td>True</td>
</tr>
<tr>
<td>The implementation of the difference scheme ( \partial_t^+ U^+_j + a\partial_x^- U^-_j = 0 ) requires the solution of linear systems of equations in every time step</td>
<td>False</td>
</tr>
<tr>
<td>The CFL condition is a necessary and sufficient condition for stability of a finite difference scheme</td>
<td>True</td>
</tr>
</tbody>
</table>
A.1.2 Heat Equation

Exercise A.1.11 Let \( u \in C^2([0, T] \times [\alpha, \beta]) \) solve the heat equation \( \partial_t u - \kappa \partial_x^2 u = 0 \). Show that for appropriate \( \tau, L, x_0 > 0 \), the function \( \tilde{u}(s, y) = u(\tau s, Ly + x_0) \) solves \( \partial_s \tilde{u} - \partial_y^2 \tilde{u} = 0 \) in \( (0, T') \times (0, 1) \).

Exercise A.1.12 Derive a mathematical model for a diffusion process that includes sinks and sources of the diffusing substance, described by a function \( f \in C([0, T] \times [0, 1]) \).

Exercise A.1.13 Let \( u \in C^2([0, T] \times [0, 1]) \) solve the heat equation \( \partial_t u - \partial_x^2 u = 0 \) with homogeneous Dirichlet boundary conditions. Prove that

\[
\frac{d}{dt} \frac{1}{2} \int_0^1 (\partial_x u(t, x))^2 \, dx \leq 0
\]

and deduce the uniqueness of solutions for the heat equation with general Dirichlet boundary conditions.

Exercise A.1.14 The construction of a solution via a separation of variables consists in finding functions \( u_n(t, x) = v_n(t)w_n(x) \) that solve the heat equation and the prescribed boundary conditions. A solution of the initial boundary value problem is then obtained by determining coefficients \( \alpha_n \) such that

\[
u(t, x) = \sum_{n=1}^{\infty} \alpha_n v_n(t)w_n(x)
\]

converges in an appropriate sense and satisfies \( u(0, x) = u_0(x) \).

(i) Construct pairs \( (v_n, w_n) \) such that \( u_n(t, x) = v_n(t)w_n(x) \) satisfies \( \partial_t u_n - \partial_x^2 u_n = 0 \) in \( (0, T) \times (0, 1) \) and \( u_n(t, 0) = u_n(t, 1) = 0 \) for all \( t \in (0, T) \).

(ii) Assume that the function \( u_0 \in C([0, 1]) \) is given as

\[
u_0(x) = \sum_{n=1}^{\infty} \gamma_n \sin(n\pi x).
\]

Construct the solution of the corresponding initial boundary value problem for the heat equation.

Remark It can be shown that every function \( u_0 \in C([0, 1]) \) can be represented in the specified form.

Exercise A.1.15

(i) Show that the explicit Euler scheme is unstable if \( \Delta t > \Delta x^2 / 2 \) by constructing appropriate initial data.
(ii) Show that numerical solutions obtained with the Crank–Nicolson scheme do in general not satisfy a discrete maximum principle.

**Exercise A.1.16** Show that the discretization in the space of the heat equation leads to a stiff initial value problem \( \frac{\partial_t U}{C} + AU = 0 \), \( U(0) = U_0 \), which admits a unique solution on every time interval \([0, T]\).

**Exercise A.1.17** For \( a, b \in \mathbb{R} \) and \( n \in \mathbb{N} \), let \( A \in \mathbb{R}^{n \times n} \) be the bandmatrix

\[
A = \begin{bmatrix}
 a & b \\
 b & \ddots & \ddots \\
 \ddots & \ddots & b \\
 & \ddots & \ddots & b \\
 b & a
\end{bmatrix}.
\]

Show that \( A \) has the eigenvalues \( \lambda_p = a + 2b \cos(p\pi/(n + 1)), p = 1, 2, \ldots, n \).

**Hint:** Show that for \( a = 0 \), corresponding eigenvectors \( v_p \in \mathbb{R}^n \) are given by \( v_{p,j} = \sin(pj\pi/(n + 1)), j = 1, 2, \ldots, n \).

**Exercise A.1.18** Let \( J \in \mathbb{N} \) and set \( \Delta x = 1/J \).

(i) Prove that the vectors \( \varphi_p \in \mathbb{R}^{J+1}, p = 1, \ldots, J - 1 \), given by \( \varphi_{p,j} = \sqrt{2} \sin(pj\Delta x), j = 0, 1, \ldots, J \), define an orthonormal basis for \( \ell^2_{0,\Delta x} = \{ V \in \mathbb{R}^{J+1} : V_0 = V_{J+1} = 0 \} \) with respect to the inner product

\[
(V, W)_{\Delta x} = \Delta x \sum_{j=0}^{J} V_j W_j.
\]

(ii) Show that the vectors \( \varphi_p \) are eigenvectors of the operator \( -\partial^+_x \partial^-_x : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1} \), defined by

\[
(-\partial^+_x \partial^-_x) V_j = \begin{cases} 
0 & \text{for } j = 0, J, \\
-\partial^+_x \partial^-_x V_j & \text{for } j = 1, 2, \ldots, J - 1.
\end{cases}
\]

**Hint:** Use that \( \varphi_{p,j} = \sqrt{2} \text{Im}(\omega^{pj}) \) with \( \omega = e^{j\pi \Delta x} \).

**Exercise A.1.19**

(i) Show formally that the function

\[
u(t, x) = \frac{1}{(4\pi t)^{1/2}} \int_{\mathbb{R}} e^{-|x-y|^2/(4t)} u_0(y) \, dy
\]

solves the heat equation \( \partial_t u - \partial^2_x u = 0 \) in \((0, T) \times \mathbb{R}\) for every \( T > 0 \).
(ii) Explain why we can expect that \( u(t,x) \to u_0(x) \) as \( t \to 0 \), e.g., for piecewise constant initial data \( u_0 \) and \( x = 0 \).

(iii) Let \( u_0(x) = 1 \) for \( x \geq 0 \) and \( u_0(x) = 0 \) for \( x < 0 \). Show that \( u(t,x) \) is positive for all \( t \in (0,T) \) and \( x \in \mathbb{R} \), and conclude that information is propagated with infinite speed.

(iv) How does the formula in (i) have to be modified to provide a solution of the heat equation \( \partial_t u - \kappa \partial_x^2 u = 0 \)?

Exercise A.1.20

(i) Show that the \( \theta \)-method is well defined for every choice of \( \theta \) and every choice of \( \Delta t, \Delta x > 0 \).

(ii) Show that the \( \theta \)-method is unstable if \( \theta < 1/2 \) and \( \lambda = \Delta t/\Delta x^2 > 1/2 \).

Quiz A.1.2 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

| The larger the constant \( \kappa > 0 \) in the heat equation \( \partial_t u - \kappa \partial_x^2 u = 0 \), the faster is the diffusion process |
| The explicit Euler scheme is stable if \( \Delta x/\Delta t \leq 1/2 \) |
| The \( \theta \)-method is explicit for \( \theta < 1/2 \) and implicit for \( \theta \geq 1/2 \) |
| The implicit Euler scheme requires the solution of a linear system of equations in every time step, whose system matrix is diagonally dominant and irreducible |
| The Crank–Nicolson scheme approximates the exact solution of the heat equation with an error of order \( \mathcal{O}(\Delta t^2 + \Delta x^2) \) if \( u \in C^3([0,T] \times [0,1]) \) |

A.1.3 Wave Equation

Exercise A.1.21

(i) Determine functions \( u_n(t,x) = v_n(t)w_n(x) \), \( n \in \mathbb{N} \), that satisfy the wave equation in \((0,T) \times (0,1)\) subject to homogeneous Dirichlet boundary conditions.

(ii) Assume that \( u_0, v_0 \in C([0,1]) \) satisfy

\[
 u_0(x) = \sum_{n \in \mathbb{N}} \alpha_n \sin(n\pi x), \quad v_0(x) = \sum_{n \in \mathbb{N}} \beta_n \sin(n\pi x)
\]

with given sequences \((a_n)_{n \in \mathbb{N}}, (b_n)_{n \in \mathbb{N}}\). Derive a representation formula for the solution of the wave equation \( \partial_t^2 u - c^2 \partial_x^2 u = 0 \) in \((0,T) \times (0,1)\) with homogeneous Dirichlet boundary conditions and initial conditions \( u(0,x) = u_0(x) \) and \( \partial_t u(0,x) = v_0(x) \) for all \( x \in [0,1] \).
Exercise A.1.22  Let \( u \in C^2([0, T] \times \mathbb{R}) \) solve the wave equation \( \frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0 \) with initial conditions \( u(0, x) = u_0(x) \) and \( \partial_t u(0, x) = v_0(x) \) for all \( x \in \mathbb{R} \).

(i) By introducing the variables \( \xi = x + ct \) and \( \eta = x - ct \), show that the function \( \tilde{u}(\xi, \eta) = u(t, x) \) satisfies \( \frac{\partial \tilde{u}}{\partial \xi} \tilde{u} + \partial_{\eta} \tilde{u} = 0 \) and deduce that \( \tilde{u}(\xi, \eta) = f(\xi) + g(\eta) \).

(ii) Conclude that there exist functions \( f, g \in C(\mathbb{R}) \) such that \( u(t, x) = f(x + ct) + g(x - ct) \).

(iii) Determine \( f \) and \( g \) in terms of \( u_0 \) and \( v_0 \).

Exercise A.1.23

(i) We consider the wave equation \( \frac{\partial^2 u}{\partial t^2} - c^2 \frac{\partial^2 u}{\partial x^2} = 0 \) in \( (0, T) \times \mathbb{R}_{>0} \) subject to the boundary condition \( u(t, 0) = 0 \) for all \( t \in (0, T) \). Use d’Alembert’s formula to represent the solution for the initial conditions \( u(0, x) = u_0(x) \) and \( \partial_t u(0, x) = v_0(x) \) for \( x \in \mathbb{R}_{>0} \), where \( u_0(x) = 0 \).

(ii) Specify the solution for \( u_0(x) = \max\{0, 1 - |x - 2|\} \) and sketch it for \( t = 0, 5/6, 2, 7/3, 15/6 \).

Exercise A.1.24

(i) Prove the energy conservation principle for the wave equation with homogeneous Neumann boundary conditions.

(ii) Deduce uniqueness of solutions for solutions of the wave equation with homogeneous Dirichlet or homogeneous Neumann boundary conditions.

Exercise A.1.25  Show that the explicit finite difference scheme for the wave equation is unstable if \( \mu = c \Delta t / \Delta x > 1 \).

Exercise A.1.26  Let \( (\xi_k)_{k \in \mathbb{N}_0} \) be a sequence of real numbers that for \( \alpha, \beta \in \mathbb{R} \) and all \( k \in \mathbb{N} \) satisfies the recursion

\[
\begin{bmatrix}
\xi_k \\
\xi_{k+1}
\end{bmatrix} = A \begin{bmatrix}
\xi_{k-1} \\
\xi_k
\end{bmatrix}, \quad A = \begin{bmatrix}
0 & 1 \\
\alpha & \beta
\end{bmatrix}.
\]

(i) Show that if the eigenvalues \( \lambda_1, \lambda_2 \in \mathbb{C} \) of \( A \) satisfy \( |\lambda_i| < 1, i = 1, 2 \), then there exists \( c > 0 \) such that \( |\xi_k| \leq c \) for all \( k \in \mathbb{N}_0 \).

(ii) Show that if the eigenvalues \( \lambda_1, \lambda_2 \in \mathbb{C} \) of \( A \) coincide and satisfy \( |\lambda_i| = 1, i = 1, 2 \), or if \( \max_{i=1,2} |\lambda_i| > 1 \), then there exist unbounded sequences \( (\xi_k)_{k \in \mathbb{N}_0} \) that satisfy the recursion.

Exercise A.1.27  For \( J \in \mathbb{N} \), let \( \Delta x = 1/J \) and let \( V, W \in \mathbb{R}^{J+1} \).

(i) Prove the discrete product rule

\[
\partial_x^- (W_j V_j) = W_j (\partial_x^- V_j) + (\partial_x^+ W_{j-1}) V_{j-1}.
\]
(ii) Deduce the summation-by-parts formula

\[
\Delta x \sum_{j=0}^{J-1} (\partial_x^+ W_j) V_j = -\Delta x \sum_{j=1}^{J} W_j (\partial_x^- V_j) + W_J V_J - W_0 V_0,
\]

and explain its relation to the integration-by-parts formula.

**Exercise A.1.28** Let \( J \in \mathbb{N}, \Delta x = 1/J, \) and \((U_j^k) \in \mathbb{R}^{(K+1) \times (J+1)}\). Prove the identities

\[
(\partial_t^+ \partial_t^- U_j^k)(\widehat{\partial_t} U_j^k) = \frac{1}{2} \partial_t^- (U_j^k)^2,
\]

and

\[
\frac{c^2}{4} \partial_x^+ \partial_x^- (U_j^{k+1} + 2U_j^k + U_j^{k-1})(\widehat{\partial_t} U_j^k)
\]

\[
= \frac{c^2}{2\Delta t} (\partial_x^+ \partial_x^- (U_j^{k+1/2} + U_j^{k-1/2}))(U_j^{k+1/2} - U_j^{k-1/2}),
\]

for \( 1 \leq j \leq J - 1 \) and \( 1 \leq k \leq K - 1 \), where \( U_j^{k\pm 1/2} = (U_j^k + U_j^{k\pm 1})/2 \).

**Exercise A.1.29**

(i) Prove that the implicit difference scheme for the wave equation is well defined, i.e., leads to regular linear systems of equations in all time steps.

(ii) Show that the implicit difference scheme for the wave equation has a consistency error \( \mathcal{O}(\Delta t^2 + \Delta x^2) \).

**Exercise A.1.30** The wave equation \( \partial_t^2 u = \partial_x^2 u \) can be written as the system \( \partial_t u = v, \partial_t v = \partial_x^2 u \). Discretize this system with backward difference quotients in time and a central difference quotient in space and analyze the stability of the resulting scheme.

**Quiz A.1.3** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>The total kinetic energy of a solution for the wave equation is constant</td>
<td>True/False</td>
</tr>
<tr>
<td>The explicit difference scheme for the wave equation is stable if ( c \Delta t \leq \Delta x )</td>
<td>True/False</td>
</tr>
<tr>
<td>The implicit scheme for the wave equation unconditionally satisfies a discrete maximum principle</td>
<td>True/False</td>
</tr>
<tr>
<td>The larger the constant ( c ) in the wave equation ( \partial_t^2 u - c^2 \partial_x^2 u = 0 ), the smaller is the wave speed</td>
<td>True/False</td>
</tr>
<tr>
<td>The discretization of the initial condition ( \partial_t u(0, x) = v_0(x) ) with a central difference quotient leads to a consistency error ( \mathcal{O}(\Delta t^2) )</td>
<td>True/False</td>
</tr>
</tbody>
</table>
**A.1.4 Poisson Equation**

**Exercise A.1.31** Let $\Omega = (0, 1)^2$ and let $f \in C(\overline{\Omega})$ be given by

$$f(x_1, x_2) = \sum_{m,n \in \mathbb{N}} \alpha_{m,n} \sin(m\pi x_1) \sin(n\pi x_2).$$

Compute $-\Delta u_{m,n}$ for $u_{m,n}(x_1, x_2) = \sin(\pi m x_1) \sin(\pi n x_2)$ and construct the solution of the Poisson problem $-\Delta u = f$ in $\Omega$ and $u = 0$ on $\partial \Omega$.

**Exercise A.1.32**

(i) Show that it is sufficient to assume that $-\Delta u \leq 0$ to prove the maximum principle $\max_{x_2 \in \overline{\Omega}} u(x) \leq \max_{x \in \partial \Omega} u(x)$.

(ii) Let $u \in C^2(\overline{\Omega})$ solve $-\Delta u = f$ in $\Omega$ and $u = 0$ on $\partial \Omega$. Apply the maximum principle to an appropriately defined function $v = u + \|f\|_{C(\overline{\Omega})} w$ to prove that

$$\|u\|_{C(\overline{\Omega})} \leq \max_{x \in \partial \Omega} \frac{|x|^2}{2d} \|f\|_{C(\overline{\Omega})}.$$ 

Is it possible to improve this estimate?

**Exercise A.1.33**

(i) Use Gauss’s theorem to show that for $u, v \in C^2(\overline{\Omega})$, we have

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

$$\int_{\Omega} (u \Delta v - v \Delta u) \, dx = \int_{\partial \Omega} (u \nabla v \cdot n - v \nabla u \cdot n) \, ds.$$

(ii) Let $u_1, u_2 \in C^2(\overline{\Omega})$ be solutions of the boundary value problem $-\Delta u = f$ in $\Omega$ and $u = 0$ on $\partial \Omega$. Show that

$$\int_{\Omega} |\nabla (u_1 - u_2)|^2 \, dx = 0$$

and conclude that $u_1 = u_2$.

**Exercise A.1.34** Let $x_0 \in \mathbb{R}^d$ for $d \in \{2, 3\}$, $a > 0$, and $u \in C^1(\overline{B_a(x_0)})$.

(i) Show that in polar coordinates with respect to $x_0$, we have

$$\nabla u \cdot n = \partial_r u$$

on $\partial B_{a'}(x_0)$ for every $0 < a' \leq a$. 

(ii) Show that
\[
\lim_{r \to 0} \frac{1}{|\partial B_r(x_0)|} \int_{\partial B_r(x_0)} u(s) \, ds = u(x_0),
\]
where $|\partial B_r(x_0)|$ denotes the surface measure of $\partial B_r(x_0)$.

**Exercise A.1.35** Let $w : \mathbb{R}^2 \to \mathbb{R}$ be a quadratic polynomial and $\Delta x = 1/J$ for some $J \in \mathbb{N}$. For $j, m \in \mathbb{Z}^2$, let $x_{j,m} = (j, m)\Delta x$ and $W_{j,m} = w(x_{j,m})$. Show that
\[
\Delta_h W_{j,m} = \partial_{x_1}^+ \partial_{x_1}^- W_{j,m} + \partial_{x_2}^+ \partial_{x_2}^- W_{j,m} = \Delta w(x_{j,m})
\]
for all $j, m \in \mathbb{Z}^2$.

**Exercise A.1.36** Let $J \in \mathbb{N}$, set $L = (J-1)^2$, and let $X \in \mathbb{R}^{(J-1) \times (J-1)}$ and $A \in \mathbb{R}^{L \times L}$ be defined by
\[
X = \begin{bmatrix}
2 & -1 & & & \\
-1 & \ddots & \ddots & & \\
& \ddots & \ddots & -1 & \\
& & 1 & 2 & \\
\end{bmatrix}, \quad A = \begin{bmatrix}
X & -I \\
-I & \ddots & \ddots & \\
& \ddots & \ddots & -I \\
& & -I & X \\
\end{bmatrix},
\]
where $I \in \mathbb{R}^{(J-1) \times (J-1)}$ denotes the identity matrix. Show that $A$ is diagonally dominant and irreducible.

**Exercise A.1.37** Let $AU = F$ be the linear system of equations corresponding to the discretized Poisson problem $-\Delta u = f$ in $\Omega = (0, 1)^2$ with homogeneous Dirichlet boundary conditions. Show that the Richardson scheme for the iterative solution of the linear system can be identified with an explicit discretization of the heat equation.

**Exercise A.1.38** Let $A \in \mathbb{R}^{n \times n}$ be the system matrix corresponding to the discretization of the Poisson problem. Use the discrete maximum principle to show that for the matrix $B = A^{-1}$, we have $b_{ij} \geq 0$, $i, j = 1, 2, \ldots, n$.

**Exercise A.1.39** We consider a swimming pool that has the horizontal shape of an annulus and assume that the stationary temperature distribution is independent of the vertical direction. Moreover, we assume that the temperature is prescribed at the boundary. We thus consider the two-dimensional boundary value problem
\[
-\Delta u = 0 \text{ in } \Omega = B_{r_2}(0) \setminus \overline{B_{r_1}(0)}, \quad u = u_1 \text{ on } \partial B_{r_1}(0), \quad u = u_2 \text{ on } \partial B_{r_2}(0)
\]
for given real numbers $0 < r_1 < r_2$ and $u_1, u_2 \in \mathbb{R}$.
(i) Show that for \( g \in C^2(\mathbb{R}) \) and \( r(x_1, x_2) = (x_1^2 + x_2^2)^{1/2} \), we have
\[
\Delta(g \circ r) = g''(r) + r^{-1}g'(r) = r^{-1}(rg'(r))'.
\]

(ii) Justify the assumption \( u = \hat{u} \circ r \) and solve the Poisson problem for the swimming pool with \( r_1 = 10, r_2 = 20, \) and \( u_1 = 20, u_2 = 40 \).

(iii) On which radius do you have to swim to be surrounded by water of 30°C?

**Exercise A.1.40**

(i) Let \( L > 1 \) and for \( \alpha_\ell, p_\ell \in \mathbb{R}, \) \( 0 \leq \ell \leq L \), assume that \( \alpha_\ell < 0 \) for \( \ell = 1, 2, \ldots, L \), and
\[
\sum_{\ell=0}^{L} \alpha_\ell \geq 0, \quad \sum_{\ell=0}^{L} \alpha_\ell p_\ell \leq 0.
\]
Suppose further that \( p_0 \geq 0 \) or \( \sum_{\ell=0}^{L} \alpha_\ell = 0 \). Show that \( p_0 \geq \max_{1 \leq \ell \leq L} p_\ell \) implies \( p_0 = p_1 = \cdots = p_L \).

(ii) Let \( (U_{j,m})_{0 \leq j, m \leq J} \) be the finite difference approximation of the Poisson problem \(-\Delta u = f\) in \( \Omega = (0, 1)^2 \) and \( u = u_D \) on \( \partial \Omega \). Assume that \( f \leq 0 \) and show that
\[
\max_{1 \leq j, m \leq J-1} U_{j,m} \leq \max_{x_j, m \in \partial \Omega} u_D(x_j, m).
\]

**Quiz A.1.4** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>If ( f ) is constant, then the solution of the Poisson problem (-\Delta u = f) in ( \Omega ), ( u</td>
<td>_{\partial \Omega} = 0 ), is constant.</td>
</tr>
<tr>
<td>If ( u_1 ) and ( u_2 ) are harmonic functions, then ( u_1 - u_2 ) is also a harmonic function.</td>
<td>True</td>
</tr>
<tr>
<td>The finite difference discretization of the Poisson problem has a consistency error of order ( O(\Delta^2) ).</td>
<td>True</td>
</tr>
<tr>
<td>If (-\Delta u \geq 0), then ( \max_{x \in \Omega} u(x) \geq \max_{x \in \partial \Omega} u(x) ).</td>
<td>True</td>
</tr>
<tr>
<td>If ( U_i, i = 1, 2 ), are the coefficient vectors of finite difference solutions of Poisson problems (-\Delta U_i = F_i) with homogeneous Dirichlet boundary conditions, then we have ( |U_1 - U_2|<em>{\infty} \leq c|F_1 - F_2|</em>{\infty} ) with a constant ( c &gt; 0 ).</td>
<td>True</td>
</tr>
</tbody>
</table>

**A.1.5 General Concepts**

**Exercise A.1.41** Write the initial boundary value problem for the wave equation as an abstract boundary value problem \( F(u) = 0 \) in \( U \) and \( G(u) = 0 \) on \( \partial U \) by defining appropriate mappings \( F \) and \( G \).
Exercise A.1.42 Discuss the well-posedness of the heat equation with homogeneous boundary conditions. In particular, discuss the effect of perturbations of initial data.

Exercise A.1.43

(i) Let \( A, b, c \in \mathbb{R}^n \) and \( f \in C(U) \). Assume that \( u \in C^2(U) \) satisfies

\[
\sum_{i,j=1}^{n} a_{ij} \frac{\partial^2}{\partial z_i \partial z_j} u(z) + \sum_{j=1}^{n} b_j \frac{\partial}{\partial z_j} u(z) + c u(z) = f(z)
\]

for all \( z \in U \). Suppose that \( A = Q^T \Lambda Q \) is diagonalizable and define \( \tilde{u}(\xi) = u(Q\xi) \). Determine the partial differential equation satisfied by \( \tilde{u} \).

(ii) Determine the type of the following partial differential equations:

\[
\partial_t u + \Delta u = f \quad \text{in} \quad (0, T) \times \Omega \subset \mathbb{R}_{\geq 0} \times \mathbb{R}^d,
\]

\[
\partial_{x_1}^2 u - 3 \partial_{x_1} \partial_{x_2} u + \partial_{x_2}^2 u = 0 \quad \text{in} \quad \Omega \subset \mathbb{R}^2,
\]

\[
\partial_t u - \partial_{x_1}^2 u + \partial_{x_2} u = f \quad \text{in} \quad (0, T) \times \Omega \subset \mathbb{R}_{\geq 0} \times \mathbb{R}^2.
\]

Exercise A.1.44 Formulate the stability and consistency of the difference scheme for the transport equation in an abstract framework and apply the Lax–Richtmyer theorem to derive an error estimate.

Exercise A.1.45 Formulate the stability and consistency of the implicit difference scheme for the wave equation in an abstract framework and apply the Lax–Richtmyer theorem to derive an error estimate.

Exercise A.1.46 Let \( L_h(U_h) = \ell_h \) be the linear system of equations resulting from the discretization of a boundary value problem. Here, boundary nodes have been eliminated appropriately. Suppose that the discretization is stable in the sense that

\[
\| U_h \|_{\ell,N_h} \leq c \| \ell_h \|_{r,N_h}
\]

for all \( \ell_h \in \mathbb{R}^{N_h} \) and \( U_h \in \mathbb{R}^{N_h} \) with \( L_h(U_h) = \ell_h \) and with norms \( \| \cdot \|_{\ell,N_h} \) and \( \| \cdot \|_{r,N_h} \) on \( \mathbb{R}^{N_h} \). Show that \( L_h : \mathbb{R}^{N_h} \to \mathbb{R}^{N_h} \) is an isomorphism and conclude that the discrete problem admits a unique solution.

Exercise A.1.47 For \( u \in C^2([0, 1]^2) \) and grid points \( x_{j,m} = (j, m)\Delta x, \ 0 \leq j, m \leq J \), with \( \Delta x = 1/J \), define the interpolant of \( u \) by

\[
\mathcal{I}_h u = (u(x_{j,m}))_{0 \leq j, m \leq J} \in \mathbb{R}^{(J+1)^2}.
\]

Show that with the norm \( \| V \|_{\infty} = \max_{0 \leq j, m \leq J} |V_{j,m}| \) on \( \mathbb{R}^{(J+1)^2} \), we have for \( \Delta x \to 0 \) that

\[
\| \mathcal{I}_h u \|_{\infty} \to \| u \|_{C([0,1]^2)}.
\]
Exercise A.1.48 Discuss the discretization and numerical solution of the three-dimensional Poisson problem \(-\Delta u = f\) in \(\Omega = (0, 1)^3\) and \(u = 0\) on \(\partial \Omega\). Provide a stability estimate, determine the consistency error, and specify the resulting linear system of equations.

Exercise A.1.49

(i) Let \(J \in \mathbb{N}\) and \(\Delta x = 1/J\). Let \((\varphi_p : p = 1, \ldots, J - 1)\) be the eigenvectors of \(-\partial_{x_1}^+ \partial_{x_1}^-\) given by \(\varphi_{p,j} = \sqrt{2} \sin(pj\pi \Delta x), 0 \leq j \leq J\). Show that the vectors \(\psi_{(p,q)} \in \mathbb{R}^{(J+1)^2}\), defined by

\[
\psi_{(p,q),(j,m)} = \varphi_{p,j} \varphi_{q,m} = 2 \sin(pj\pi \Delta x) \sin(qm\pi \Delta x)
\]

are eigenvectors of the operator \(-\Delta_h = -\partial_{x_1}^+ \partial_{x_1}^- - \partial_{x_2}^+ \partial_{x_2}^-\) and that they define an orthonormal basis of the space of grid functions with vanishing boundary conditions with respect to the inner product

\[
(V, W)_{\Delta x} = \Delta x^2 \sum_{j,m=0}^{J} V_{j,m} W_{j,m}.
\]

(ii) Carry out a stability analysis of the \(\theta\)-method for approximating the two-dimensional heat equation.

Exercise A.1.50 Assume that \(F_h(U_h) = L_h U_h - \ell_h\) is a discretization of a linear boundary value problem which is bounded and convergent in the sense that we have the estimates

\[
\|LV_h\|_{r,N_h} \leq c_1 \|V_h\|_{r,N_h}, \quad \|S_h u - U_h\|_{r,N_h} \leq c_2 h^\alpha \|u\|_{C^{k+s}(\bar{U})},
\]

for every \(V_h \in \mathbb{R}^{N_h}\) and continuous and discrete solutions \(u \in C^{k+s}(\bar{U})\) and \(U_h \in \mathbb{R}^{N_h}\). Show that the numerical scheme is consistent of order \(\alpha\).

Quiz A.1.5 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True or False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every initial value problem (y'(t) = f(t, y(t))) for (t \in (0, T]), (y(0) = y_0), with a continuous function (f) defines a well-posed boundary value problem</td>
<td>True</td>
</tr>
<tr>
<td>The implicit Euler scheme for the two-dimensional heat equation is unconditionally stable</td>
<td>True</td>
</tr>
<tr>
<td>Every partial differential equation admits solutions</td>
<td>True</td>
</tr>
<tr>
<td>The implicit scheme for the two-dimensional wave equation is stable if (\Delta t \leq c \Delta x)</td>
<td>True</td>
</tr>
<tr>
<td>The equation (\partial_t^2 u - 4 \partial_x u + \partial_x^2 u = 0) is elliptic</td>
<td>False</td>
</tr>
</tbody>
</table>
A.1.6 Projects

Project A.1.1

(i) Numerically solve the transport equation $\partial_t u + \partial_x u = 0$ in $(0, T) \times (0, 1)$ for $T = 1$ with boundary condition $u(t, 0) = 0$ and initial condition defined by $u_0(x) = 1$ for $0.4 \leq x \leq 0.6$ and $u_0(x) = 0$ otherwise, using a forward difference quotient in time and a backward difference quotient in space. Try the pairs of discretization parameters

$$(\Delta t, \Delta x) = \frac{1}{80}(2, 2), \quad (\Delta t, \Delta x) = \frac{1}{80}(2, 1), \quad (\Delta t, \Delta x) = \frac{1}{80}(1, 2).$$

Check for which of the pairs the CFL condition is satisfied, and compare the numerical solution with the exact solution of the transport equation.

(ii) Modify your code to obtain an approximation scheme for the equation $\partial_t u + a(x) \partial_x u = 0$,

where $a(x) > 0$ is a given function. How should the CFL condition be formulated for nonconstant functions $a$? Test your code with $a(x) = (1 + 4x^2)^{1/2}$ and initial conditions $u_0(x) = 1$ if $0.05 \leq x \leq 0.25$ and $u_0(x) = 0$ otherwise. Compare the numerical solutions for various discretization parameters.

(iii) Run your program with $a(x) = -1$ and initial condition $u_0(x) = 1$ for $0.4 \leq x \leq 0.6$ and $u_0(x) = 0$ otherwise. Are there pairs of discretization parameters so that the CFL condition is satisfied?

(iv) Modify your code so that a forward difference is realized. Here the boundary condition $u(1, t) = 0$ for $t \in [0, T]$ is given. Derive the CFL condition for this approximation scheme and try different choices for $\Delta t$ and $\Delta x$.

Project A.1.2 An upwinding scheme for the transport equation is defined by

$$U_j^{k+1} = \begin{cases} (1 - \mu_j^k)U_j^k + \mu_j^k U_{j-1}^k, & \mu_j^k \geq 0, \\ (1 + \mu_j^k)U_j^k - \mu_j^k U_{j+1}^k, & \mu_j^k < 0, \end{cases}$$

where $\mu_j^k = a(t_k, x_j) \Delta t / \Delta x$. Implement the scheme and test it with different initial conditions, discretization parameters, the function $a(x) = \sin(x)$, and boundary conditions defined by $u(0, t) = u(1, t) = 0$. Discuss your results and the validity of a CFL condition.

Project A.1.3

(i) Implement a $\theta$-midpoint scheme to approximately solve the initial boundary value problem $\partial_t u = \kappa \partial_x^2 u$ in $(0, T) \times (0, 1)$ for $T = 1$ and $\kappa = 1/100$, $u(0, x) = \sin \pi x$ for $x \in (0, 1)$, and $u(t, 0) = u(t, 1) = 0$ for $t \in [0, T]$. Set
\( \Delta x = 0.05 \) and experimentally determine \( \Delta t \) so that the scheme is stable for \( \theta = 0 \).

(ii) Verify that the exact solution of the problem is given by

\[
u(t, x) = \sin(\pi x) \exp \left( -\kappa \pi^2 t \right).
\]

For \( \theta = 1/2, \theta = 3/4, \) and \( \theta = 1 \), determine the approximation error using the displaying format `long` at the point \((t, x) = (1, 0.5)\) for \( \Delta x = \Delta t = 2^{-j}/10 \), for \( j = 2, 3, \ldots, 5 \). Plot the errors in one figure using the commands `semilogy` and `hold on/off`. What is your conclusion?

(iii) Modify your code to allow for a right-hand side \( f \), i.e., the partial differential equation \( \partial_t u - \kappa \partial_x^2 u = f \), and solve the initial boundary value problem in \((0, T) \times (0, 1)\) with \( T = 2, f(x) = (x-1/2)^2 \), homogeneous Dirichlet boundary conditions at \( x = 0 \) and \( x = 1 \), and the initial condition defined by \( u_0(x) = 1 \) if \( 0.45 \leq x \leq 0.55 \). Compare the numerical solutions for various discretization parameters and \( \theta = 0, 1/2, 1 \).

**Project A.1.4**

(i) Numerically solve the wave equation \( \partial_t^2 u - \partial_x^2 u = 0 \) in \((0, T) \times (0, 1)\) with homogeneous Dirichlet boundary conditions and initial conditions \( v_0(x) = 0 \) and \( u_0(x) = \sin(\pi x) \), using an explicit difference scheme with discretization parameters

\[
(\Delta t, \Delta x) = \frac{1}{40}(2, 2), \quad (\Delta t, \Delta x) = \frac{1}{40}(2, 1), \quad (\Delta t, \Delta x) = \frac{1}{40}(1, 2).
\]

Compare your results and explain differences in the numerical solutions.

(ii) Change the initial conditions to

\[
u_0(x) = 0, \quad v_0(x) = \begin{cases} 1 & \text{if } 0.4 \leq x \leq 0.6, \\ 0 & \text{otherwise}, \end{cases}
\]

and run the program for different pairs of discretization parameters.

(iii) Experimentally investigate the violation of a discrete energy conservation principle by plotting the quantity

\[
\Gamma^k = \frac{\Delta x}{2} \sum_{j=1}^{J-1} |\partial_x^+ U_j^k|^2 + \frac{\Delta x}{2} \sum_{j=1}^{J} |\partial_x^- U_j^k|
\]

as functions of \( k = 0, 1, \ldots, K - 1 \).

**Project A.1.5** The sound of a stringed instrument is defined by the occurrence of different overtones. To verify experimentally that the wave equation captures this effect, we consider a string of length \( \ell > 0 \) that is plucked at time \( t = 0 \) at a
Fig. A.1  Initial
displacement of a string

position \( x_p \in (0, \ell) \) by a distance \( H > 0 \), so that we have

\[
\begin{align*}
  u_0(x) &= \begin{cases} 
    Hx/x_p & \text{for } x \leq x_p, \\
    H(\ell - x)/(\ell - x_p) & \text{for } x \geq x_p,
  \end{cases}
\end{align*}
\]

cf. Fig. A.1. We assume that the tone is sampled at a position \( x_s \), e.g., by a hole of
an acoustic or a pickup in the case of an electric instrument. The initial velocity is
assumed to vanish, and the ends of the string are fixed. Assuming for simplicity that
\( \ell = 1 \), a separation of variables in the wave equation implies that we have

\[
u(t, x) = \sum_{m=1}^{\infty} \beta_m \cos(\omega_m t) \sin(m \pi x)
\]

with \( \omega_m = m \pi c \) and \( c = (\varrho/\sigma)^{1/2} \). Numerically solve the wave equation with
\( c = 2, T = 2, x_p = 1/8 \), and \( H = 1/100 \), and use your approximations to determine
coefficients \( \alpha_m, m = 1, 2, \ldots, K \), such that

\[
U_{j_s}^k = \sum_{m=1}^{K} \alpha_m \cos(\omega_m t_k),
\]

where \( j_s \) is the index corresponding to the grid-point which equals \( x_s = 1/4 \). Plot
the harmonics \( w_m(t) = \alpha_m \cos(\omega_m t), m = 1, 2, \ldots, 6 \), as functions of \( t \in [0, T] \),
and visualize the amplitude distribution by plotting the function \( m \mapsto |\alpha_m| \). Try
other values for \( x_p \) and \( x_s \) and compare the results. Is it necessary to solve the wave
equation in order to determine the coefficients \( \alpha_m \)?

**Project A.1.6** Implement the unconditionally stable implicit scheme

\[
\partial_t^+ \partial_x^- U_j^n = \frac{1}{4} \partial_x^+ \partial_x^- (U_{j+1}^{n+1} + 2U_j^n + U_{j-1}^{n-1})
\]

for approximating the wave equation \( \partial_t^2 u = \partial_x^2 u \) in \( (0, T) \times (0, 1) \) with homogeneous
Dirichlet boundary conditions and initial conditions \( u(x, 0) = u_0(x) \) and \( \partial_t u(0, x) = v_0(x) \) for \( x \in (0, 1) \) and given functions \( u_0, v_0 \in C([0, 1]) \). Use a discretization of
the initial condition \( \partial_t u(0, x) \) that leads to quadratic convergence. Test your program
with the exact solution

\[
u(t, x) = \cos(\pi t) \sin(\pi x).
\]
Project A.1.7 Define functions \( f, g, u_D \) so that \( u(x, y) = \sin(\pi x) \sin(\pi y) \) is the solution of the boundary value problem
\[
-\Delta u = f \quad \text{in } \Omega = (0, 1)^2, \\
u = u_D \quad \text{on } \Gamma_D = [0, 1] \times \{0\}, \\
\partial_n u = g \quad \text{on } \Gamma_N = \partial \Omega \setminus \Gamma_D.
\]
Introduce ghost points and use centered difference quotients to approximate the normal derivative on \( \Gamma_N \). Experimentally verify that the scheme is quadratically convergent.

Project A.1.8 We consider an oven occupying the region \( \Omega = (0, 0.4) \times (0, 0.3) \times (-\ell_z, \ell_z) \) and assume that the back of the oven is constantly heated to a temperature of \( \theta = 200^\circ \text{C} \), the front is either open or closed, and all other sides are thermally insulated, i.e., that \( \partial_n \theta = 0 \). When the front side is open, we assume that \( \theta = 20^\circ \text{C} \), and when it is closed we have \( \partial_n \theta = 0 \), cf. Fig. A.2. At time \( t = 0 \) we assume that the temperature inside the oven is uniformly given by \( \theta = 200^\circ \text{C} \). A mathematical model is obtained from the physical laws that heat density is proportional to temperature, i.e., \( w = \varrho c_p \theta \), heat flux is proportional to the temperature gradient, i.e., \( q = -\kappa \nabla \theta \), and thermal energy is conserved, i.e., \( \partial_t w + \text{div} \, q = 0 \). In particular, we use the density \( \varrho = 1.435 \cdot 10^{-3} \text{kg/m}^3 \), the heat conductivity \( \kappa = 0.024 \text{W/m K} \), and the heat capacity \( c_p = 1.007 \cdot 10^3 \text{J/kg K} \). To reduce the dimension of the problem, we replace \( \theta \) by its horizontal average, i.e., we consider
\[
\theta'(t, x_1, x_2) = \frac{1}{2\ell_z} \int_{-\ell_z}^{\ell_z} \theta(t, x_1, x_2, x_3) \, dx_3.
\]
Formulate an initial boundary value problem to describe the averaged temperature \( \theta' \) in \( \Omega' = (0, 0.4) \times (0, 0.3) \). Implement a Crank–Nicolson scheme and simulate the following scenarios: (i) oven open for 30s, closed for 30s, open for 30s; (ii) oven closed for 30s, open for 60s. Decide on the basis of your simulations whether it is energetically preferable to open the oven once for a long period or twice for shorter periods. Discuss limitations of the model and the numerical method.

Fig. A.2 Schematic description of the cross-section of an oven.
A.2 Elliptic Partial Differential Equations

A.2.1 Weak Formulation

Exercise A.2.1 For $\Omega \subset \mathbb{R}^2$ and $u \in C^2(\Omega)$, let $\tilde{u}(r, \phi) = u(r \cos \phi, r \sin \phi)$.

(i) Show that

$$
\nabla u(r \cos \phi, r \sin \phi) = \begin{bmatrix}
\partial_r \tilde{u}(r, \phi), & r^{-1} \partial_\phi \tilde{u}(r, \phi)
\end{bmatrix}^T
$$

and

$$
\Delta u(r \cos \phi, r \sin \phi) = \partial_r^2 \tilde{u}(r, \phi) + r^{-1} \partial_\phi \tilde{u}(r, \phi) + r^{-2} \partial_\phi^2 \tilde{u}(r, \phi).
$$

(ii) Verify that the function $\tilde{u}(r, \phi) = r^{\pi/\alpha} \sin(\phi \pi / \alpha)$ is harmonic.

Exercise A.2.2 For an open set $U \subset \mathbb{C}$, let $f : U \to \mathbb{C}$ be complex differentiable, i.e., for every $z \in U$ there exists $f'(z_0) \in \mathbb{C}$ such that

$$
\lim_{h \to 0} \frac{f(z_0 + h) - f(z_0)}{h} = f'(z_0),
$$

where $h \to 0$ represents an arbitrary sequence of complex numbers that converges to zero. Show that the functions $u, v : U \to \mathbb{R}$ defined by $f(x + iy) = u(x, y) + iv(x, y)$ satisfy the equations

$$
\partial_x u = \partial_y v, \quad \partial_y u = -\partial_x v
$$
in $U$ and that they are harmonic, i.e., satisfy $-\Delta u = 0$ and $-\Delta v = 0$ in $U$.

Exercise A.2.3 Assume that $\Omega$ is connected and $\Gamma \neq \emptyset$. Prove that there exists at most one solution of the weak formulation of the Poisson problem.

Exercise A.2.4

(i) Show that the function

$$
\phi(x) = \begin{cases}
0 & \text{for } |x| \geq 1, \\
e^{-1/(1-|x|^2)} & \text{for } |x| < 1
\end{cases}
$$

satisfies $\phi \in C^\infty(\mathbb{R}^d)$. Is $\phi$ an analytic function?

(ii) Let $h \in C(\Omega)$ and assume that

$$
\int_{\Omega} h v \, dx = 0
$$

for all $v \in C^\infty(\Omega)$ with $v = 0$ on $\partial \Omega$. Prove that $h = 0$ in $\Omega$. 
Exercise A.2.5 Sketch the nonclassical solution \( u(r, \phi) = r^{\pi/\alpha} \sin(\phi \pi/\alpha) \) of the Poisson problem and its gradient for \( \alpha \in \{\pi/2, \pi, 3\pi/2\} \).

Exercise A.2.6 Show that the Dirichlet energy is convex, i.e., \( I((1 - t)u + tv) \leq (1 - t)I(u) + tI(v) \) for \( t \in [0, 1] \).

Exercise A.2.7 Show that the functional
\[
I(v) = \int_1^1 x^2(u'(x))^2 \, dx
\]
has no minimizer in \( C^1((-1, 1)) \) subject to the boundary conditions \( v(-1) = -1 \) and \( v(1) = 1 \).

Exercise A.2.8 Let \( \Omega \subset \mathbb{R}^d \) be open, bounded, and connected, and let \( I_D \subset \partial \Omega \) be nonempty. Prove that
\[
\|v\| = \left( \int_{\Omega} |\nabla v|^2 \, dx \right)^{1/2}
\]
defines a norm on \( V = \{v \in C^1(\overline{\Omega}) : v|_{I_D} = 0\} \).

Exercise A.2.9 Let \( A \in \mathbb{R}^{n \times n} \) and \( b \in \mathbb{R}^n \).

(i) Show that \( x \in \mathbb{R}^n \) satisfies \( Ax = b \) if and only if
\[
(Ax) \cdot y = b \cdot y
\]
for all \( y \in \mathbb{R}^n \).

(ii) Assume that \( A \) is symmetric and positive definite. Show that there exists a matrix \( B \in \mathbb{R}^{n \times n} \) such that the unique solution of the linear system \( Ax = b \) is the unique minimizer of the mapping
\[
z \mapsto \frac{1}{2} |Bz|^2 - b \cdot z.
\]

Exercise A.2.10 Let \( F : \mathbb{R}^n \to \mathbb{R} \) be such that there exist constants \( c_1, c_2 > 0 \) so that
\[
F(z) \geq c_1|z| - c_2.
\]
Assume that \( F \) is lower semicontinuous, i.e., whenever \( z_j \to z \) as \( j \to \infty \) then \( F(z) \leq \lim \inf_{j \to \infty} F(z_j) \). Show that \( F \) has a global minimizer and provide an example of a function \( F \) that satisfies the conditions but is not continuous.
**Quiz A.2.1** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>If $f \in C^1(\Omega)$, $\Gamma_D = \partial\Omega$, and $u_D = 0$, then the Poisson problem has a classical solution</td>
<td>True</td>
<td>The weak formulation of the Poisson problem specifies a function $u \in C^1(\Omega) \cap C^2(\Omega)$, and the system of equations $Ax = b$ can be identified with a formulation $a(x, y) = b(y)$ for all $y \in \mathbb{R}^n$.</td>
</tr>
<tr>
<td>Every bilinear form $a : V \times V \rightarrow \mathbb{R}$ is symmetric</td>
<td>True</td>
<td>Every nonnegative functional $I : V \rightarrow \mathbb{R}$ on a Banach space $V$ has a minimizer</td>
</tr>
</tbody>
</table>

**A.2.2 Elementary Functional Analysis**

**Exercise A.2.11** Let $a : V \times V \rightarrow \mathbb{R}$ be symmetric, bilinear, and positive semidefinite. Prove that

$$a(v, w) \leq (a(v, v))^{1/2} (a(w, w))^{1/2}.$$  

**Exercise A.2.12** Let $I \subset \mathbb{R}$ be a closed interval. Show that $C^1(I)$ is complete with respect to the norm

$$\|v\| = \sup_{x \in I} |v(x)| + \sup_{x \in I} |v'(x)|$$  

but not with respect to the norm

$$\|v\| = \int_I |v(x)| + |v'(x)| \, dx.$$  

**Exercise A.2.13** Let $V$ be a Banach space and let $a : V \times V \rightarrow \mathbb{R}$ be bilinear, symmetric, and positive semidefinite. Moreover, assume that there exist $c_1, c_2 > 0$ such that

$$c_1 \|v\|_V \leq (a(v, v))^{1/2} \leq c_2 \|v\|_V$$

for all $v \in V$. Show that $a$ defines a scalar product on $V$ and that $V$ is a Hilbert space with this scalar product.

**Exercise A.2.14** Let $V, W$ be $n$- and $m$-dimensional linear spaces. Use the Riesz representation theorem to prove that $L(V, W)$ is isomorphic to $\mathbb{R}^{n \times m}$, i.e., that linear mappings can be identified with matrices.
Exercise A.2.15 Prove that the set of square summable sequences \( \ell^2(\mathbb{N}) = \{(v_j)_{j \in \mathbb{N}} : \sum_{j \in \mathbb{N}} v_j^2 < \infty\} \) is a Hilbert space.

Exercise A.2.16

(i) Show that the linear operator \( A : V \to W \) is continuous if and only if it is bounded in the sense that there exists \( c > 0 \) such that

\[
\|Av\|_W \leq c\|v\|_V
\]

for all \( v \in V \).

(ii) Let \( A : V \to W \) be linear and bounded and let \( \|A\|_{L(V,W)} \) be the infimum of all such constants \( c > 0 \). Show that for all \( v \in V \) we have

\[
\|Av\|_W \leq \|A\|_{L(V,W)}\|v\|_V.
\]

(iii) Show that \( A \mapsto \|A\|_{L(V,W)} \) defines a norm on the space of linear and bounded operators \( L(V, W) \) such that it is a Banach space.

Exercise A.2.17 Determine all matrices \( M \in \mathbb{R}^{n \times n} \) such that the bilinear mapping \( a : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \)

\[
a(x, y) = x^T My,
\]

satisfies the conditions of the (i) Riesz representation theorem and (ii) Lax–Milgram lemma.

Exercise A.2.18 Let \( (v_j)_{j \in \mathbb{N}} \subset \ell^2(\mathbb{N}) \) be defined by \( v_{j,n} = \delta_{j,n} \), i.e.,

\[
v_j = [0, \ldots, 0, 1, 0, \ldots].
\]

Prove that the sequence converges weakly and determine the weak limit.

Exercise A.2.19 Let \( (v_j)_{j \in \mathbb{N}} \) be a weakly convergent sequence in the Banach space \( V \). Show that the sequence is bounded, that the weak limit is unique, and that the weak limit coincides with the strong limit provided it exists.

Exercise A.2.20

(i) Assume that \( A : V \to W \) is a linear and compact operator between Banach spaces \( V \) and \( W \). Show that \( A \) is bounded.

(ii) Assume that \( A : V \to \mathbb{R}^n \) is linear and bounded. Show that \( A \) is compact.
Quiz A.2.2  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every finite-dimensional subspace of a Banach space is closed</td>
<td>True</td>
</tr>
<tr>
<td>The scalar product of a Hilbert space defines a symmetric and positive definite bilinear form</td>
<td>True</td>
</tr>
<tr>
<td>Every bounded sequence in a Banach space has a convergent subsequence</td>
<td>True</td>
</tr>
<tr>
<td>The space $C([0,1])$ equipped with the maximum norm is a Banach space</td>
<td>True</td>
</tr>
<tr>
<td>Every linear operator $A : V \to W$ between finite-dimensional spaces is bounded</td>
<td>True</td>
</tr>
</tbody>
</table>

A.2.3  Sobolev Spaces

Exercise A.2.21

(i) Prove that for $1 < p, q < \infty$ with $1/p + 1/q = 1$ and all $a, b \in \mathbb{R}_{\geq 0}$ we have

$$ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q.$$  

(ii) Prove Hölder’s inequality

$$\int_{\Omega} |uv| \, dx \leq \|u\|_{L^p(\Omega)} \|v\|_{L^q(\Omega)}$$

for $u \in L^p(\Omega)$ and $v \in L^q(\Omega)$ with $1/p + 1/q = 1$.

*Hint:* Consider the case $\|u\|_{L^p(\Omega)} = \|v\|_{L^q(\Omega)} = 1$ first.

Exercise A.2.22  Prove that $|a + b|^p \leq |a + b|^{p-1} (|a| + |b|)$ for all $a, b \in \mathbb{R}$ and use Hölder’s inequality to deduce Minkowski’s inequality

$$\|u + v\|_{L^p(\Omega)} \leq \|u\|_{L^p(\Omega)} + \|v\|_{L^p(\Omega)}$$

for all $u, v \in L^p(\Omega)$.

Exercise A.2.23

(i) Let $\Omega \subset \mathbb{R}^d$ such that $\partial \Omega$ is piecewise of class $C^1$. Show that Gauss’s theorem is equivalent to the identity

$$\int_{\Omega} u(\partial_i v) \, dx = -\int_{\Omega} (\partial_i u) v \, dx + \int_{\partial \Omega} un_i \, ds$$

for all $u, v \in C^1(\overline{\Omega})$ and $i = 1, 2, \ldots, d$. 
(ii) Let $\Omega \subset \mathbb{R}^d$ be open. Show that for all $u \in C^1(\mathbb{R}^d)$ and $\phi \in C_0^\infty(\mathbb{R}^d)$, we have

$$\int_{\Omega} u(\partial_i \phi) \, dx = -\int_{\Omega} (\partial_i u) \phi \, dx$$

for $i = 1, 2, \ldots, d$.

**Exercise A.2.24** Let $d \in \mathbb{N}$, $s \in \mathbb{R}$, and $\Omega = B_1(0) \subset \mathbb{R}^d$, and define $u(x) = |x|^s$ for $x \in \Omega \setminus \{0\}$.

(i) Determine all $s \in \mathbb{R}$ so that $u \in L^p(\Omega)$.

(ii) Determine all $s \in \mathbb{R}$ so that $u \in W^{1,p}(\Omega)$.

**Exercise A.2.25** Let $(\Omega_j)_{j=1}^J$ be an open partition of $\Omega$, i.e., the $\overline{\Omega_j} = \Omega_1 \cup \cdots \cup \Omega_J$, $\Omega_j$ is open for $j = 1, 2, \ldots, J$, and $\Omega_j \cap \Omega_\ell = \emptyset$ for $j \neq \ell$. Let $u \in C(\overline{\Omega})$ be such that $u|_{\Omega_j} \in C^1(\overline{\Omega_j})$ for $j = 1, 2, \ldots, J$. Show that $u$ is weakly differentiable and $u \in W^{1,p}(\Omega)$ for all $1 \leq p \leq \infty$.

**Exercise A.2.26** Let $\Omega \subset \mathbb{R}^d$ be bounded, open, and connected, and assume that $u \in W^{1,p}(\Omega)$ satisfies $\nabla u = 0$. Use a convolution kernel and the identity $\partial_i J_\varepsilon \ast \phi = J_\varepsilon \ast \partial_i \phi$ to prove that $u$ is constant.

**Exercise A.2.27** Let $u, v \in W^{1,2}(\Omega)$. Prove that $uv \in W^{1,1}(\Omega)$ with $\nabla (uv) = u\nabla v + v\nabla u$.

*Hint:* Approximate $u$ and $v$ by smooth functions.

**Exercise A.2.28** Let $\Omega = (0, 1)^2$ and $\Gamma_{B1} = [0, 1] \times \{0\}$. Prove that for every $u \in C^\infty([0, 1]^2)$ with $u|_{\Gamma_{B1}} = 0$, we have

$$\|u\|_{L^p(\Gamma_{B1})} \leq c \| \partial_2 u \|_{L^p(\Omega)}.$$  

**Exercise A.2.29** Let $\Omega = B_{1/2}(0) \subset \mathbb{R}^2$ and define $u(x) = \log(\log(|x|))$. Show that $u \in W^{1,2}(\Omega)$ but $u \not\in L^\infty(\Omega)$ and $u \not\in C(\Omega)$.

*Hint:* Use that for $F(r) = |\log(r)|^{-1}$, we have $F'(r) = 1/\log^2(r)$.

**Exercise A.2.30**

(i) Let Lip$(\Omega)$ be the set of all Lipschitz continuous functions on $\Omega$ and define

$$\|u\|_{\text{Lip}(\Omega)} = \|u\|_{L^\infty(\Omega)} + \sup_{x \neq y} \frac{|u(x) - u(y)|}{|x - y|}$$

for $u \in \text{Lip}(\Omega)$. Show that Lip$(\Omega)$ is a Banach space.

(ii) Use the Arzelà–Ascoli theorem to show that the embedding Lip$(\Omega) \rightarrow C(\overline{\Omega})$ is compact.
Quiz A.2.3  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

- Functions in $L^\infty(\Omega)$ can be approximated by continuous functions
- Every differentiable function is weakly differentiable
- The weak partial derivative $\partial_i$ defines a bounded linear operator $\partial_i: W^{1,2}(\Omega) \to L^2(\Omega)$
- If $p > q$ and $\Omega$ is bounded, then $W^{k,q}(\Omega) \subset W^{k,p}(\Omega)$
- If $\Omega \subset \mathbb{R}^3$ is a bounded Lipschitz domain, then functions in $W^{1,p}(\Omega)$, $p \geq 3/2$, are continuous

A.2.4 Weak Solutions

Exercise A.2.31  For $\alpha \in (0, 2\pi)$, let

$$\Omega = \{ r(\cos \phi, \sin \phi) : 0 < r < 1, 0 < \phi < \alpha \},$$

$\Gamma_D = \partial \Omega$, and $\Gamma_N = \emptyset$, and define $f = 0$ in $\Omega$, and

$$u_D(r, \phi) = \begin{cases} 0 & \text{for } \phi \in \{0, \alpha\}, \\ \sin(\phi \pi/\alpha) & \text{for } r = 1. \end{cases}$$

Prove that

$$u(r, \phi) = r^{\pi/\alpha} \sin(\phi \pi/\alpha)$$

is a weak solution of the Poisson problem.

Exercise A.2.32

(i) Derive a weak formulation for the boundary value problem

$$\begin{cases} -\text{div}(K \nabla u) + b \cdot \nabla u + cu = f & \text{in } \Omega, \\ u = u_D & \text{on } \Gamma_D, \\ (K \nabla u) \cdot n = g & \text{on } \Gamma_N. \end{cases}$$

(ii) Specify conditions on the coefficients that lead to the existence of a unique weak solution $u \in H^1(\Omega)$. 
**Exercise A.2.33** Let $\Omega = (0, 1)^2$ and define for $j, k = 1, 2, \ldots, N$,

$$\phi_{j,k}(x_1, x_2) = \sin(\pi x_1 j / N) \sin(\pi x_2 k / N)$$

and let $V_h$ be the span of $(\phi_{j,k} : j, k = 1, 2, \ldots, N)$. Compute the stiffness matrix for the bilinear mapping related to the Laplace operator.

**Exercise A.2.34** Prove by constructing an appropriate example that the Poisson problem is in general not $H^2$-regular if the domain is not convex.

**Exercise A.2.35** Let $(u_j) \subset C^3(\overline{\Omega}) \cap H^1_0(\Omega)$ be such that $u_j \rightharpoonup u$ in $H^1(\Omega)$. Assume that for a sequence $(f_j)_{j \in \mathbb{N}} \subset L^2(\Omega)$, we have $f_j \rightharpoonup f$ in $L^2(\Omega)$ and

$$\int_{\Omega} \nabla u_j \cdot \nabla v \, dx = \int_{\Omega} f_j v \, dx$$

for all $j \in \mathbb{N}$ and all $v \in H^1_0(\Omega)$. Assume further that

$$\int_{\Omega} |D^2 u_j|^2 \, dx \leq \int_{\Omega} |f_j|^2 \, dx$$

for all $j \in \mathbb{N}$. Show that $u \in H^2(\Omega)$ with $\|D^2 u\|_{L^2(\Omega)} \leq \|f\|_{L^2(\Omega)}$.

**Exercise A.2.36** Let $u \in C^3(\Omega)$. Show that we have

$$|\Delta u|^2 - |D^2 u|^2 = \text{div} \left( \nabla u \Delta u - \frac{1}{2} \nabla |\nabla u|^2 \right).$$

**Exercise A.2.37** Let $Q' \subset \mathbb{R}^{d-1}$ be open and let $h \in C^2(Q')$ be concave. Prove that for all $x' \in Q'$ we have

$$\sum_{i=1}^{d-1} \partial_{i}^2 h(x') \leq 0.$$

**Exercise A.2.38** For $f \in L^2(\Omega)$ let $u_f \in H^1_0(\Omega)$ be the unique solution of the Poisson problem with the right-hand side $f$, and let $L_f : L^2(\Omega) \to L^2(\Omega)$ be defined by $f \mapsto u_f$. Verify and prove whether the operator $L_f$ is linear, bounded, injective, surjective, and compact.

**Exercise A.2.39** Show that the boundary value problem

$$\Delta^2 u = f \text{ in } \Omega, \quad u = \Delta u = 0 \text{ on } \partial \Omega$$

has a unique weak solution $u \in H^2(\Omega)$. 
Exercise A.2.40  Show that the Neumann problem

\[-\Delta u = f \text{ in } \Omega, \quad \partial_n u = g \text{ on } \Gamma_N = \partial \Omega\]

has a unique solution \( u \in H^1(\Omega) \) satisfying \( \int_{\Omega} f \, dx = 0 \) if and only if

\[\int_{\Omega} f \, dx + \int_{\partial \Omega} g \, ds = 0.\]

Quiz A.2.4  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>The stiffness matrix of a Galerkin method for an elliptic boundary value problem is symmetric and positive definite</td>
<td>True</td>
</tr>
<tr>
<td>Elliptic partial differential equations with constant coefficients define (H^2)-regular problems</td>
<td>True</td>
</tr>
<tr>
<td>Every weak solution of an elliptic boundary value problem is a classical solution</td>
<td>True</td>
</tr>
<tr>
<td>The existence and uniqueness of a weak solution for the Poisson problem is a consequence of the Riesz representation theorem</td>
<td>True</td>
</tr>
<tr>
<td>The Galerkin approximation ( u_h \in V_h ) of an elliptic boundary value problem minimizes the distance to the exact solution in the set ( V_h )</td>
<td>True</td>
</tr>
</tbody>
</table>

A.2.5  Projects

Project A.2.1  Compute approximate solutions \( u_m \in \mathcal{P}_m[0,1] \) of the one-dimensional Poisson problem \(-u'' = f \) in \( \Omega = (0, 1) \) with boundary conditions \( u(0) = u(1) = 1 \), by numerically solving the system of equations

\[-u''(x_i) = f(x_i), \quad i = 1, 2, \ldots, m - 1, \quad u_m(x_0) = u_m(x_m) = 0,\]

where \( x_i = i/m \) for \( i = 0, 1, \ldots, m \). Test the method for the right-hand sides \( f(x) = 1 \) and \( f(x) = \text{sign}(x - 1/2) \). Investigate the decay of the error \( \max_{i=0,\ldots,m} |u(x_i) - u_m(x_i)| \) for \( m \to \infty \) and the conditioning of the linear system of equations.

Project A.2.2  For \( m \in \mathbb{N}_0 \) and \( j = 1, 2 \), consider the subspaces \( V_m^{(j)} \subset H^1_0(\Omega) \) for \( \Omega = (0, 1) \) defined by

\[V_m^{(1)} = \left\{ \sum_{0 \leq j + k \leq m} \alpha_{j,k} x^j (1 - x)^k \right\}, \quad V_m^{(2)} = \left\{ \sum_{0 \leq j \leq m} \beta_j \sin(\pi jx) \right\}.

Compute the Galerkin approximations of the Poisson problem $-u'' = 1$ in $\Omega = (0, 1)$ with Dirichlet boundary conditions $u(0) = u(1) = 0$. Comparatively, investigate the convergence of the methods and the properties of the linear systems of equations.

**Project A.2.3** We consider the square $Q = [0, 1]^2$ and the triangle $T = \{(x_1, x_2) \in \mathbb{R}^2 : 0 \leq x_2 \leq x_1 \leq 1\}$, and the mapping

$$\Phi : Q \to T, \quad (\xi_1, \xi_2) \mapsto (\xi_1, \xi_1 \xi_2).$$

Explain the identity

$$\int_T f(x) \, dx = \int_0^1 \int_0^1 f(\xi_1, \xi_1 \xi_2) \xi_1 \, d\xi_2 \, d\xi_1$$

and use it to define a quadrature rule on $T$. Determine experimentally, for which $s \in \mathbb{R}$, the function $x \mapsto |x|^s$ is integrable on $T$.

**Project A.2.4** Use MATLAB routines for the approximate solution of ordinary differential equations to numerically determine the level set $f(x_1, x_2) = f(y)$ of a given function $f : \mathbb{R}^2 \to \mathbb{R}$ and a point $y \in \mathbb{R}^2$. Visualize the graph of the function, its gradient via arrows, and some of its level sets for the cases $f(x) = |x|^2$ and $f(x_1, x_2) = \sin(\pi x_1) \cos(\pi x_2)$.

**Project A.2.5** We consider the convolution kernel $J : \mathbb{R}^2 \to \mathbb{R}$ defined by

$$J(x) = \begin{cases} c_2 e^{-1/(1-|x|^2)}, & |x| < 1, \\ 0, & |x| \geq 1. \end{cases}$$

Plot the functions $J_\varepsilon(x) = \varepsilon^{-d} J(x/\varepsilon)$ for $\varepsilon = 10^{-j}, j = 0, 1, 2$, using $c_2 = 1$. Use an iterated trapezoidal rule to determine $c_2$ such that $\|J\|_{L^1(\mathbb{R}^2)} = 1$. Approximate and visualize the regularizations $f_\varepsilon = J_\varepsilon * f$ for the cases $f(x_1, x_2) = |(x_1, x_2)|$ and $f(x_1, x_2) = \text{sign}(x_1) \text{sign}(x_2)$ with $\varepsilon = 10^{-j}, j = 0, 1, 2$.

**Project A.2.6** Use an iterated trapezoidal rule to approximate the integral

$$\int_{(-1,1)^2} \frac{[x_1, x_2]^T}{(x_1^2 + x_2^2)^{1/2}} \cdot \begin{bmatrix} (1 - x_1^2)(1 - x_2^2) \\ \cos(\pi x_1/2) \cos(\pi x_2/2) \end{bmatrix} \, dx_1 \, dx_2.$$

Try to improve the approximation by using integration-by-parts.

**Project A.2.7** We define the two-dimensional torus $T_{r,R}$ for radii $r, R > 0$ as the image of the mapping

$$f : [0, 2\pi]^2 \to \mathbb{R}^3, \quad (\theta, \phi) \mapsto ((R + r \cos \theta) \cos \phi, (R + r \cos \theta) \sin \phi, r \sin \theta).$$
Use the transformation formula and an iterated trapezoidal rule to approximate the surface integral of the function \( u(x, y, z) = x^2 y^3 z^4 \) on \( T_{R, r} \) for \( R = 1 \) and \( r = 1/8 \). Visualize the torus and the function \( u \) by partitioning the parameter domain into triangles and using the MATLAB command \texttt{trisurf}.

**Project A.2.8** Numerically compute the \( W^{1,2} \) norm of the functions \( u(x) = \log \log |x| \) and \( u(x) = |x|^{1/2} \) in the domain \( \Omega = (-1/2, 1/2)^2 \). Visualize the functions and their gradients.

### A.3 Finite Element Method

#### A.3.1 Interpolation with Finite Elements

**Exercise A.3.1** Prove that the interpolant \( \mathcal{I}_T v \) associated with a finite element \((T, \mathcal{P}, \mathcal{K})\) is well defined for all \( v \in W^{m,p}(T) \).

**Exercise A.3.2** For a triangle \( T \subset \mathbb{R}^2 \) with vertices \( z_0, z_1, z_2 \in \mathbb{R}^2 \), let \( z_3, z_4, z_5 \in \mathbb{R}^2 \) be the midpoints of the sides of \( T \).

(i) Show that \((T, \mathcal{P}_2(T), \mathcal{K})\) with \( \mathcal{K} = \{ \chi_j : j = 0, 1, \ldots, 5 \} \) for \( \chi_j(\phi) = \phi(z_j) \), \( j = 0, 1, \ldots, 5 \), is a finite element.

(ii) Construct the dual basis for the finite element \((T, \mathcal{P}_2(T), \mathcal{K})\).

**Exercise A.3.3** Let \( w = (w_1, w_2, \ldots, w_d) : \mathbb{R}^d \to \mathbb{R}^d \) be a polynomial vector field of degree \( m - 1 \) on \( \mathbb{R}^d \), and assume that \( w = \nabla v \) for some function \( v \in C^1(\mathbb{R}^d) \). Show that \( v \) is a polynomial of degree \( m \).

**Exercise A.3.4** Let \( \omega \subset \mathbb{R}^d \) be a bounded Lipschitz domain. Provide a constructive proof for the existence of a constant \( c_P > 0 \), such that for all \( v \in H^1(\omega) \) with

\[
\int_\omega v \, dx = 0
\]

we have \( \| v \|_{L^2(\omega)} \leq c_P \| \nabla v \|_{L^2(\omega)} \).

*Hint:* Use the mean-value theorem to represent \( v(x) \) by an integral over \( \omega \).

**Exercise A.3.5** Let \( k \in \mathbb{N} \), and define \( N = |\{ \alpha \in \mathbb{N}^d_0 : |\alpha| \leq k \}| \). Show that

\[
\mathcal{P}_k(T) = \left\{ \sum_{\alpha \in \mathbb{N}_0^d, |\alpha| \leq k} a_\alpha x^\alpha : a_\alpha \in \mathbb{R} \right\},
\]

and that the mapping

\[
\mathcal{P}_{m-1}(T) \to \mathbb{R}^N, \quad q \mapsto \left( \int_T \partial^\alpha q(x) \, dx \right)_{\alpha \in \mathbb{N}_0^d, |\alpha| \leq m-1}
\]

is an isomorphism.
Exercise A.3.6 Let $\Phi_T : \widehat{T} \to T$ be an affine diffeomorphism. Show that
\[
D\Phi_T^{-1} = (D\Phi_T)^{-1}
\]
and that both matrices are independent of $x \in T$ and $\hat{x} \in \widehat{T}$.

Exercise A.3.7 Let $(\widehat{T}, \widehat{\mathcal{P}}, \widehat{\mathcal{H}})$ be a finite element, and $\Phi_T : \widehat{T} \to T$ an affine diffeomorphism.

(i) Show that the triple $(T, \mathcal{P}, \mathcal{H})$ defined by
\[
T = \Phi_T(\widehat{T}), \quad \mathcal{P} = \{\hat{q} \circ \Phi_T^{-1} : \hat{q} \in \widehat{\mathcal{P}}\}, \quad \mathcal{H} = \{\hat{\chi} \circ \Phi_T^{-1} : \hat{\chi} \in \widehat{\mathcal{H}}\}
\]
is a finite element.

(ii) Show that for the interpolants $I_T$ and $I_{\hat{T}}$ of the finite elements $(\widehat{T}, \widehat{\mathcal{P}}, \widehat{\mathcal{H}})$ and $(T, \mathcal{P}, \mathcal{H})$, we have $(I_T v) \circ \Phi_T = I_{\hat{T}} \hat{v}$.

Exercise A.3.8 Assume that the sequence of triangulations $(\mathcal{T}_h)_{h>0}$ of the domain $\Omega \subset \mathbb{R}^2$ satisfies a minimum angle condition, i.e., we have
\[
\inf_{h>0} \min_{T \in \mathcal{T}_h} \min_{j=0,1,2} \alpha_j^T \geq c_0 > 0,
\]
where for a triangle $T \subset \mathbb{R}^d$ the numbers $\alpha_0^T, \alpha_1^T, \alpha_2^T \in (0, \pi)$ are the inner angles of $T$. Prove that the sequence is uniformly shape regular.

Exercise A.3.9 Let $0 = x_0 < x_1 < \cdots < x_n = 1$ be a partition of $(0, 1)$. Show that the subordinated cubic spline space $\mathscr{S}^{3,2}(\mathcal{T}_h) \subset C^2([0, 1])$ is not an affine family.

Exercise A.3.10 Construct a sequence of approximating, shape regular triangulations of the ring $B_2(0) \setminus \overline{B_1(0)}$.

Quiz A.3.1 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>If $T \subset \mathbb{R}^2$ is a triangle, $x_0, x_1, x_2 \in T$ are distinct points, and $\chi_j(q) = q(x_j)$ for $j = 1, 2, 3$, then $(T, \mathcal{P}_1(T), {\chi_0, \chi_1, \chi_2})$ is a finite element</td>
<td>True</td>
</tr>
<tr>
<td>There exists a constant $c &gt; 0$ such that for all piecewise polynomial functions $v \in H^1(\Omega)$, we have $|\nabla v|<em>{L^2(\Omega)} \leq c|v|</em>{L^2(\Omega)}$</td>
<td>True</td>
</tr>
<tr>
<td>If $T_1$ and $T_2$ are elements in a conforming triangulation $\mathcal{T}_h$, then we have $\text{diam}(T_1) \sim \text{diam}(T_2)$</td>
<td>True</td>
</tr>
<tr>
<td>For all $v \in H^1(T)$ there exists $q \in \mathcal{P}<em>2(T)$ such that $|\nabla (v - q)|</em>{L^2(T)} \leq c |v|_{H^1(T)}$</td>
<td>True</td>
</tr>
<tr>
<td>If $\int_T v , dx = 0$ and $\int_T Dv , dx = 0$, then there exists a polynomial $q \in \mathcal{P}<em>1(T)$ with $|v - q|</em>{L^2(T)} \leq c |Dv|_{L^2(T)}$</td>
<td>True</td>
</tr>
</tbody>
</table>
\section*{A.3.2 P1-Approximation of the Poisson Problem}

**Exercise A.3.11** Let $\mathcal{T}_h$ be a triangulation of $\Omega \subset \mathbb{R}^d$ with nodes $\mathcal{N}_h$.

(i) Show that for every $z \in \mathcal{N}_h$ there exists a unique function $\varphi_z \in \mathcal{S}^1(\mathcal{T}_h)$ with $\varphi_z(y) = \delta_{zy}$ for all $y \in \mathcal{N}_h$.

(ii) Prove that the families $(\varphi_z : z \in \mathcal{N}_h)$ and $(\varphi_z : z \in \mathcal{N}_h \setminus \Gamma_D)$ define bases for the spaces $\mathcal{S}^1(\mathcal{T}_h)$ and $\mathcal{S}^1_D(\mathcal{T}_h)$.

**Exercise A.3.12** Let $a : H^1_D(\Omega) \times H^1_D(\Omega) \to \mathbb{R}$ be a symmetric and coercive bilinear form, and let $\mathcal{T}_h$ be a triangulation of $\Omega$. Let $A = (A_{zy})_{z,y \in \mathcal{N}_h \setminus \Gamma_D}$ be for $z,y \in \mathcal{N}_h \setminus \Gamma_D$ defined by

$$A_{zy} = a(\varphi_z, \varphi_y).$$

(i) Prove that $A$ is positive definite and symmetric.

(ii) Show that for the bilinear form $a$ induced by the Poisson problem, the resulting matrix $A$ is sparse, i.e., the number of nonvanishing entries in $A$ is proportional to $|\mathcal{N}_h|$. 

**Exercise A.3.13** Let $\Omega = (0,1)^2$, $\Gamma_D = \partial \Omega$, and $f \in C(\overline{\Omega})$. Let $\mathcal{T}_h$ be the triangulation of $\Omega$ consisting of halved squares of sidelengths $h = 1/n$, and with diagonals parallel to the vector $(1,1)$. Show that the P1-finite element method and the finite difference method with a five-point stencil lead to linear systems of equations with identical system matrices.

**Exercise A.3.14** Let $(\mathcal{T}_h)_{h > 0}$ be a family of triangulations of $\Omega \subset \mathbb{R}^d$ with maximal mesh-size $h \to 0$.

(i) Show that $\cup_{h > 0} \mathcal{S}^1(\mathcal{T}_h)$ is dense in $H^1(\Omega)$.

(ii) Prove that Galerkin approximations of the Poisson problem always converge to the exact solution.

**Exercise A.3.15** Let $S \in \mathcal{T}_h$ be an inner side in a triangulation $\mathcal{T}_h$ with endpoints $z,y \in \mathcal{N}_h$ and neighboring triangles $T_1, T_2 \in \mathcal{T}_h$. Let $\alpha_1$ and $\alpha_2$ be the inner angles of $T_1$ and $T_2$ opposite to $S$, respectively. Prove that

$$A_{zy} = \int_{T_1 \cup T_2} \nabla \varphi_z \cdot \nabla \varphi_y \, dx = -\frac{1}{2} (\cot \alpha_1 + \cot \alpha_2) = -\frac{1}{2} \frac{\sin(\alpha_1 + \alpha_2)}{\sin(\alpha_1) \sin(\alpha_2)}$$

and formulate precise conditions which imply $A_{zy} \leq 0$. 

Exercise A.3.16 Let \((\mathcal{T}_h)_{h>0}\) be a triangulation of \(\Omega \subset \mathbb{R}^d\), and let \(I_\Omega = \partial \Omega\). Prove that if
\[
A_{xy} = \int_\Omega \nabla \varphi_x \cdot \nabla \varphi_y \, dx \leq 0
\]
for all distinct \(z, y \in N_h \setminus \partial \Omega\), then the Galerkin approximation \(u_h \in S_0^1(\mathcal{T}_h)\) of the Poisson problem is nonnegative, whenever the right-hand side \(f\) has this property. Show that in general this is not the case.

Exercise A.3.17 Let \(T \subset \mathbb{R}^d\) be a simplex with vertices \(z_0, z_1, \ldots, z_d \in \mathbb{R}^d\).

(i) Prove that the midpoint rule
\[
Q_T(\phi) = |T|\phi(x_T), \quad x_T = (d + 1)^{-1} \sum_{j=0}^d z_j
\]
is an exact quadrature formula for \(\phi \in P_1(T)\).

(ii) Assume \(d = 2\) and define \(\kappa_1 = \kappa_2 = \kappa_3 = 1/3\), and
\[
\xi_1 = \frac{1}{6}(4z_0 + z_1 + z_2), \quad \xi_2 = \frac{1}{6}(z_0 + 4z_1 + z_2), \quad \xi_3 = \frac{1}{6}(z_0 + z_1 + 4z_2).
\]
Show that the quadrature rule \(Q_T(\phi) = \sum_{j=1}^3 |T|\kappa_j \phi(\xi_j)\) is exact for polynomials of partial degree two.

Exercise A.3.18 Let \(W = V + V_h\), and assume that \(a : W \times W \to \mathbb{R}\) is bilinear and continuous with respect to a norm \(\| \cdot \|_h\), and assume that \(a_h\) is coercive on \(V_h\). Let \(u_h \in V_h\) satisfy \(a_h(u_h, v_h) = \ell_h(v_h)\) for all \(v_h \in V_h\), and let \(u \in V\) be such that \(a(u, v) = \ell(v)\) for all \(v \in V\).

(i) Show there exists \(c > 0\) such that
\[
c^{-1} \|u - u_h\|_h \leq \inf_{v_h \in V_h} \|u - v_h\|_h + \|a_h(u, \cdot) - \ell_h\|_{V_h^\prime}.
\]

(ii) Use the estimate to control the error induced by an approximate treatment of the domain in a Poisson problem.

Exercise A.3.19 Let \((\mathcal{T}_h)_{h>0}\) be a regular family of quasiuniform triangulations of \(\Omega \subset \mathbb{R}^d\).

(i) Show that there exists \(c > 0\), such that for all \(v_h \in S^1(\mathcal{T}_h)\) we have
\[
\|\nabla v_h\|_{L^2(\Omega)} \leq c h^{-1} \|v_h\|_{L^2(\Omega)}.
\]

(ii) Show that the quasiuniformity condition cannot be omitted in general.

(iii) Show that the estimate from (i) does not hold for functions \(v \in H^1(\Omega)\).
Exercise A.3.20  Devise and analyze a $P_1$-finite element method for the approximation of the boundary value problem

$$-\Delta u + c_0 u = f \text{ in } \Omega, \quad u|_{\partial \Omega} = 0.$$  

Quiz A.3.2  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

We have $S_1(T) = \{ v_h \in C^1(\Omega) : v_h|_T \in \mathcal{P}_1(T) \text{ for all } T \in T_h \}$

The $P_1$-finite element method for the Poisson problem satisfies a discrete maximum principle

If the solution of the Poisson problem satisfies $u \in H^2(\Omega) \cap H_0^1(\Omega)$, then we have $\| u - u_h \|_{L^2(\Omega)} \leq c h^2 \| D^2 u \|_{L^2(\Omega)}$ for the Galerkin approximation $u_h \in S_1(T_h)$

For a quasiuniform triangulation $T_h$ we have $\text{diam}(T_1) \sim \text{diam}(T_2)$ for all $T_1, T_2 \in T_h$

For all $v_h \in S_1(T_h)$ we have $\| v_h \|_{L^2(\Omega)} \leq c \| \nabla v_h \|_{L^2(\Omega)}$

A.3.3  Implementation of P1- and P2-Methods

Exercise A.3.21

(i) Let $T \subset \mathbb{R}^2$ be a triangle such that two of its sides are parallel to the coordinate axes. Let $\delta_1, \delta_2$ be the lengths of these sides. Show that the components of the gradients of the nodal basis functions belong to $\{ \delta_1^{-1}, \delta_2^{-1} \}$ and that

$$\sum_{z \in \partial \Omega \cap T} \nabla \varphi_z = 0.$$  

(ii) Let $\mathcal{T}_h$ be the triangulation of $\Omega = (0,1)^2$ with $\Gamma_D = [0,1] \times \{0\}$ shown in Fig. A.3. Manually compute the coefficients and right-hand side in the linear system of equations

$$Ax = b$$

that determines the nontrivial coefficients of the $P_1$-Galerkin approximation of the Poisson problem $-\Delta u = 1$ in $\Omega$, $u|_{\Gamma_D} = 0$, and $\partial_n u|_{\Gamma_N} = 2$.

Exercise A.3.22  Define arrays $\text{edges}$, $\text{el2edges}$, $\text{Db2edges}$, and $\text{Nn2edges}$ that specify the edges in the triangulation, edges of elements, and edges on the Dirichlet and Neumann boundaries for the triangulation shown in Fig. A.3.
Exercise A.3.23 Let $T \equiv (z_0, z_1, \ldots, z_d)$ be a simplex with positively oriented vertices $z_0, z_1, \ldots, z_d \in \mathbb{R}^d$ and define

$$X_T = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_d \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)}.$$ 

Prove that $\det X_T = d! |T|$.

Exercise A.3.24 Let $T \equiv (z_0, z_1, \ldots, z_d)$ be a simplex with positively oriented vertices $z_0, z_1, \ldots, z_d \in \mathbb{R}^d$ and define

$$X_T = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ z_0 & z_1 & \ldots & z_d \end{bmatrix} \in \mathbb{R}^{(d+1) \times (d+1)}.$$ 

Prove that the gradients of the nodal basis functions on $T$ satisfy

$$\left[ \nabla \varphi_{z_0}|_T, \ldots, \nabla \varphi_{z_d}|_T \right]^T = X_T^{-1} \begin{bmatrix} 0 \\ I_d \end{bmatrix}.$$ 

Exercise A.3.25 Let $\mathcal{T}_h$ be a triangulation.

(i) Show that for a side $S = \text{conv}\{z_0, z_1, \ldots, z_{d-1}\} \in \mathcal{T}_h$, the surface area $|S|$ is given by

$$|S| = \begin{cases} 1 & \text{if } d = 1, \\ |z_1 - z_0| & \text{if } d = 2, \\ |(z_2 - z_0) \times (z_1 - z_0)|/2 & \text{if } d = 3. \end{cases}$$

(ii) Show that for $T \in \mathcal{T}_h$, $S \in \mathcal{T}_h$, and $z \in T \cap S$, we have

$$\int_T \varphi_z \, dx = \frac{|T|}{d+1}, \quad \int_S \varphi_z \, ds = \frac{|S|}{d}.$$ 

Exercise A.3.26 Let $T \subset \mathbb{R}^d$ be a triangle or tetrahedron with vertices $z_0, z_1, \ldots, z_{d+1}$, and for $j = 1, 2, \ldots, d + 1$, let $S_j \subset \partial T$ be the side of $T$ opposite to the vertex $z_j$, with outer unit normal $n_j \in \mathbb{R}^d$. Prove that for the nodal basis function
\[ \varphi_j : T \to \mathbb{R} \text{ associated with } z_j, \text{ we have} \]
\[ \nabla \varphi_j = -\varrho_j n_j, \]
where \( \varrho_j > 0 \) is the height of \( T \) with respect to \( S_j \). Show that \( \varrho_j = |S_j|/(d|T|) \).

**Exercise A.3.27** Let \( \mathcal{T}_0 \) be a triangulation of \( \Omega \subset \mathbb{R}^d \), and let \( (\mathcal{T}_j)_{j \in \mathbb{N}} \) be the sequence of triangulations obtained from \( \mathcal{T}_0 \) by \( j = 1, 2, \ldots \) red-refinements. Prove that the sequence \( (\mathcal{T}_j)_{j \in \mathbb{N}} \) is uniformly shape regular.

**Exercise A.3.28** Let \( \mathcal{T}_h \) be a triangulation of the unit square \( \Omega = (0,1)^2 \) into halved squares. Let \( A_2 \mathbb{R}^{n \times n} \) be the finite element stiffness matrix corresponding to the Galerkin approximation of the Poisson problem. Determine bounds for the bandwidth of \( A \).

**Exercise A.3.29** Prove that the isoparametric \( P^2 \)-finite element space \( \mathcal{S}^2,iso(\mathcal{T}_h) \) is a subspace of \( C(\Omega) \).

**Exercise A.3.30** Let \( w \in \mathbb{R}^3 \) and \( t \in \mathbb{R}^5 \) be defined by
\[ w = \frac{1}{2400} [155 - \sqrt{15}, 155 + \sqrt{15}, 270], \]
\[ t = \frac{1}{21} [6 - \sqrt{15}, 9 + 2\sqrt{15}, 6\sqrt{15}, 9 - 2\sqrt{15}, 7]. \]
A quadrature rule \( Q\phi = \sum_{m=1}^{M} \kappa_m \phi(\xi_m) \) on \( \widehat{T} = \text{conv}\{0,0, (1,0), (0,1)\} \) is then defined by
\[ \xi_1 = [t_1, t_1], \quad \xi_2 = [t_2, t_1], \quad \xi_3 = [t_1, t_2], \quad \xi_4 = [t_3, t_4], \]
\[ \xi_5 = [t_3, t_3], \quad \xi_6 = [t_4, t_3], \quad \xi_7 = [t_5, t_5], \]
and \( \kappa = [w_1, w_1, w_1, w_2, w_2, w_2, w_3] \). Verify that the quadrature rule is exact of degree 5.

**Quiz A.3.3** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>The ( P^1 )-finite element discretization of the Poisson problem leads to a linear system of equations with symmetric and positive definite system matrix</td>
<td>True</td>
</tr>
<tr>
<td>The ( P^1 )-finite element stiffness matrix ( A \in \mathbb{R}^{n \times n} ) has ( O(n) ) many nonvanishing entries</td>
<td>True</td>
</tr>
<tr>
<td>The computation of the stiffness matrix ( A \in \mathbb{R}^{n \times n} ) requires ( O(n) ) many arithmetic operations</td>
<td>True</td>
</tr>
<tr>
<td>The linear system of equations ( Ax = b ) of the ( P^1 )-finite element method can be solved with ( O(n) ) operations</td>
<td>True</td>
</tr>
<tr>
<td>The ( P^1 )-finite element stiffness matrix is diagonally dominant</td>
<td>False</td>
</tr>
</tbody>
</table>
### A.3.4 P1-Approximation of Evolution Equations

**Exercise A.3.31** For given \( f \in C([0, T]; L^2(\Omega)) \) and \( u_0 \in H^1_0(\Omega) \) let \( u \in C^1([0, T]; H^1_0(\Omega)) \) be a weak solution of the heat equation.

(i) Show that there exists \( c_P > 0 \) such that

\[
\sup_{t \in [0,T]} \|u(t)\|^2 + \int_0^T \|\nabla u(t)\|^2 \, dt \leq 2\|u_0\|^2 + 2c_P \int_0^T \|f(t)\|^2 \, dt.
\]

*Hint:* Use the identity \( \frac{d}{dt} \|u\|^2 = (\partial_t u, u). *)

(ii) Deduce the uniqueness of weak solutions for the heat equation.

**Exercise A.3.32** Let \( T_h \) be a triangulation with nodes \( N_h \). Define the matrices \( M \) and \( M^h \) by

\[
M_{zy} = (\varphi_z, \varphi_y), \quad M^h_{zy} = (\varphi_z, \varphi_y)_h
\]

for \( z, y \in N_h \) with the nodal basis functions \( \varphi_z \in S^1(T_h), z \in N_h \). Show that \( M^h \) is diagonal, has nonnegative entries, and that for all \( z \in N_h \) we have

\[
M^h_{zz} = \sum_{y \in N_h} M_{zy}.
\]

**Exercise A.3.33** Let \( T_h \) be a triangulation of \( \Omega \subset \mathbb{R}^2 \). Show that there exists a constant \( c > 0 \) such that

\[
|(v, w) - (v, w)_h| \leq ch^2 \|v\|_{W^{2,2}(\Omega)} \|w\|_{W^{2,2}(\Omega)}.
\]

for all \( v, w \in C^2(\overline{\Omega}) \). Discuss weaker conditions on \( v \) and \( w \) that lead to a similar estimate.

**Exercise A.3.34** Show that approximations of the heat equation obtained with the Crank–Nicolson scheme do in general not satisfy a discrete maximum principle.

**Exercise A.3.35** Let the \( H^1 \)-projection \( Q_h : H^1_0(\Omega) \to S^1_0(T_h) \) be defined by

\[
(\nabla Q_h v, \nabla w_h) = (\nabla v, \nabla w_h)
\]

for all \( w_h \in S^1_0(T_h) \).

(i) Show that for every \( v_h \in S^1_0(T_h) \) we have

\[
\|\nabla (v - Q_h v)\| \leq \|\nabla (v - v_h)\|.
\]
(ii) Assume that the Poisson problem in $\Omega$ with $\Gamma_D = \partial \Omega$ is $H^2$-regular. Prove that

$$h^{-1} \| v - Q_h v \| + \| \nabla (v - Q_h v) \| \leq c_0 h \| D^2 v \|$$

provided that $v \in H^2(\Omega)$.

**Exercise A.3.36** Let $u \in H^1_0(\Omega)$ be the weak solution of the Poisson problem $-\Delta u = f$ in $\Omega$ subject to homogeneous Dirichlet conditions on $\partial \Omega$. Let $u_h \in \mathcal{H}^1_0(\mathcal{T}_h)$ be the Galerkin approximation. Show that with the $H^1$-projection $Q_h : H^1_0(\Omega) \to \mathcal{H}^1_0(\mathcal{T}_h)$ we have

$$u_h = Q_h u.$$

**Exercise A.3.37** Let $\mathcal{T}_h$ be a triangulation of $\Omega$, and let $u_0 \in H^2(\Omega) \cap H^1_0(\Omega)$ and $f \in L^2(\Omega)$. A semidiscrete approximation of the heat equation seeks $u_h : [0, T] \to \mathcal{H}^1_0(\mathcal{T}_h)$ with $u_h(0) = \mathcal{H} u_0$ and

$$(\partial_t u_h, v_h) + (\nabla u_h, \nabla v_h) = (f, v_h)$$

for all $v_h \in \mathcal{H}^1_0(\mathcal{T}_h)$. Show that for every $T > 0$ there exists a unique solution which is bounded independently of $h$ and $T$.

**Exercise A.3.38** Let $\mathcal{T}_h$ be a triangulation of $\Omega$, and let $u_0, v_0 \in H^2(\Omega) \cap H^1_0(\Omega)$, and $f \in L^2(\Omega)$. A semi-discrete approximation of the wave equation seeks $u_h : [0, T] \to \mathcal{H}^1_0(\mathcal{T}_h)$ with $u_h(0) = \mathcal{H} u_0, \partial_t u_h(0) = \mathcal{H} v_0$, and

$$(\partial^2_t u_h, v_h) + (\nabla u_h, \nabla v_h) = (f, v_h)$$

for all $v_h \in \mathcal{H}^1_0(\mathcal{T}_h)$. Show that for every $T > 0$ there exists a unique solution which is bounded independently of $h$ and $T$.

**Exercise A.3.39** Let $(z^k_{x})_{k=0,\ldots,K} \subset \mathcal{H}^1_0(\mathcal{T}_h)$ and $(b_{x})_{k=0,\ldots,K} \subset H^1_0(\Omega)\prime$ satisfy

$$(d_t z^k_x, v_h) + \frac{1}{4} (\nabla [z^k_x + 2z^{k-1}_x + z^{k-2}_x], \nabla v_h) = b_{x}(v_h)$$

for all $v_h \in cS^1_0(\mathcal{T}_h)$. Show that for $k = 1, 2, \ldots, K$, we have

$$\| d_t z^k_h \|^2 + \frac{1}{2} \| \nabla z^{k-1/2}_h \|^2 \leq \| d_t z^1_h \|^2 + \| \nabla z^{1/2}_h \|^2 + \frac{T}{2} \sum_{k=2}^{K} \| b_{x} \|_{H^1_0(\Omega)\prime}^2.$$

**Exercise A.3.40** Show that for $T \in \mathcal{T}_h$ with $T = \text{conv}\{z_0, z_1, \ldots, z_d\}$, we have for $0 \leq m, n \leq d$ that

$$\int_{T} \varphi_{zm} \varphi_{zn} \ dx = \frac{|T|(1 + \delta_{mn})}{(d+1)(d+2)}, \quad \int_{T} \mathcal{H}[\varphi_{zm} \varphi_{zn}] \ dx = \frac{|T|\delta_{mn}}{d+1}.$$
Quiz A.3.4  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A maximum principle for the $P_1$-approximation of the heat equation holds if and only if a maximum principle for the corresponding discretization of the Poisson problem is satisfied</td>
<td></td>
</tr>
<tr>
<td>The $\theta$-scheme is stable if $(1/2 - \theta)\tau h_{\min}^2 \leq 1$ with the minimal mesh-size $h_{\min}$</td>
<td></td>
</tr>
<tr>
<td>The $\theta$-method for the approximation of the heat equation leads to a linear system of equations $AU = b$</td>
<td></td>
</tr>
<tr>
<td>If $\theta = 1$ and the Poisson problem is $H^2$-regular, then we have that $|u(t_k) - u_k| = \mathcal{O}(h^2 + \tau^2)$</td>
<td></td>
</tr>
<tr>
<td>The midpoint scheme for the wave equation is unconditionally stable</td>
<td></td>
</tr>
</tbody>
</table>

A.3.5  Projects

Project A.3.1

(i) Modify the MATLAB program that realizes the $P_1$-finite element method for the Poisson problem to approximate the boundary value problem

$$\begin{align*}
-\text{div}(K\nabla u) &= f \text{ in } \Omega, \\
 u &= u_D \text{ on } \Gamma_D, \\
(K\nabla u) \cdot n &= g \text{ on } \Gamma_N,
\end{align*}$$

where $K: \Omega \to \mathbb{R}^{d\times d}$ is a given piecewise continuous mapping such that $K(x)$ is symmetric and positive definite for almost every $x \in \Omega$. Test your code with $\Omega = (0, 1) \times (0, 2)$, $\Gamma_N = \{1\} \times (0, 2)$, $\Gamma_D = \partial\Omega \setminus \Gamma_N$, $u(x, y) = x^2y$, and

$$K(x, y) = \begin{bmatrix} 2 & \sin(x) \\ \sin(x) & 2 \end{bmatrix}.$$  

(ii) Modify the MATLAB program that realizes the $P_1$-finite element method for the Poisson problem to approximate the boundary value problem

$$\begin{align*}
-\Delta u &= f \text{ in } \Omega, \\
 u + \alpha \partial_n u &= g \text{ on } \partial\Omega.
\end{align*}$$

Test your code for $\Omega = (0, 1)^2$, $\alpha = 2$, and $u(x, y) = x^2 + y^2$.  

Project A.3.2  Let $\gamma \in (0, 2\pi]$ and define

$$\Omega_\gamma = (-1, 1)^2 \cap \{x = r(\cos \phi, \sin \phi) : r > 0, 0 < \phi < \gamma\}.$$  

Let $u_D(r, \phi) = r^{\pi/\gamma} \sin(\phi \pi / \gamma)$ for $x = r(\cos \phi, \sin \phi) \in \Gamma_D = \partial\Omega$. The exact solution of the Poisson problem with $f = 0$ is then given by $u(r, \phi) =$
\[ r^{\pi/\gamma} \sin(\phi \pi/\gamma). \] Determine the experimental convergence rates on sequences of uniform triangulations of \( \Omega \) for \( \gamma = \ell \pi/2, \ell = 1, 2, \ldots, 4 \), for the discrete errors

\[
\delta_h^{L^2} = \|u_h - J_h u\|_{L^2(\Omega)}, \quad \delta_h^{L^\infty} = \|u_h - J_h u\|_{L^\infty(\Omega)},
\]

and for the error

\[
e_h^{H^1} = \|\nabla(u_h - u)\|_{L^2(\Omega)},
\]

where the latter integral can be approximated with the midpoint rule. Discuss your results.

**Project A.3.3** The optimal constant \( c_h > 0 \) in the inverse estimate

\[
\|\nabla v_h\|_{L^2(\Omega)} \leq c_h h^{-1} \|v_h\|_{L^2(\Omega)}
\]

for all \( v_h \in \mathcal{R}_h \) can be obtained via an eigenvalue problem and the Rayleigh quotient

\[
R(v_h) = \frac{\|\nabla v_h\|_{L^2(\Omega)}^2}{\|v_h\|_{L^2(\Omega)}^2}.
\]

Use the power method to find approximations of \( c_h \) for different sequences of uniformly refined triangulations. Choose a function \( v \in H^1(\Omega) \) and use its nodal interpolants on a sequence of uniformly refined triangulations to illustrate experimentally that an estimate of the form \( \|\nabla v\|_{L^2(\Omega)} \leq c \|v\|_{L^2(\Omega)} \) fails in general.

**Project A.3.4** For \( \varepsilon > 0 \) define the triangles \( T = \text{conv}\{(-1,0), (1,0), (0,\varepsilon)\} \) and \( T^\pm = \text{conv}\{(\pm 1,0), (0,0), (0,\varepsilon)\} \), and the triangulations \( \mathcal{R}_h^1 = \{T\} \) and \( \mathcal{R}_h^2 = \{T^+, T^-\} \) of the same domain \( \Omega_\varepsilon \). Compute the interpolation errors

\[
\|\nabla(u - J_h u)\|_{L^2(\Omega)}
\]

for the function \( u(x,y) = 1 - x^2 \) for \( \varepsilon = 10^{-j}, j = 1, 2, \ldots, 5 \), and comment on the relevance of the minimum angle condition.

**Project A.3.5** Use a sequence of approximating triangulations of the unit disk \( \Omega = B_1(0) \) to approximately solve the Poisson problem

\[-\Delta u = 1 \quad \text{in} \ \Omega, \quad u = 0 \quad \text{on} \ \Gamma_D = \partial \Omega,
\]

whose exact solution is given by \( u(x) = (|x|^2 - 1)/4 \). and determine the experimental convergence rates in \( L^2(\Omega) \) and \( H^1(\Omega) \). Repeat the experiment imposing the Neumann boundary condition \( \partial_n u = 1/2 \) on half of the boundary.
Project A.3.6 Write a C-program `nodal_basis.c` that computes for a given triangulation of a two- or three-dimensional domain $\Omega \subset \mathbb{R}^d$ specified by arrays $\mathcal{C} \mathcal{N}$ and $\mathcal{N} \mathcal{E}$ the volumes and midpoints of elements, and the elementwise gradients of the nodal basis functions. Use the formula
\[
\nabla \varphi_j |_T = -\frac{|S_z|}{d(T)} n_{S_z},
\]
for an element $T$ and a vertex $z \in T$, with opposite side $S_z \subset \partial T$, and outer unit normal $n_{S_z}$ on $S_z$. Use the MATLAB interface MEX to call the routine within MATLAB, and test it for triangulations consisting of two elements.

Project A.3.7 Implement the midpoint scheme for approximating the wave equation allowing for Neumann boundary conditions. Test your implementation for meaningful initial and boundary conditions, and experimentally verify an energy conservation principle. Augment the code allowing for a damping term in the wave equation, i.e., approximating the equation $\partial^2_t u - \alpha \Delta \partial_t u - c^2 \Delta u = f$, and test the method with an exact solution.

Project A.3.8 We consider a can in a refrigerator and want to determine the time needed to cool the liquid inside the can below a given temperature. We assume that the metal surface of the can immediately and throughout has the same temperature as its environment inside the refrigerator. To derive a model that describes the temperature changes inside the can, we use the physical laws that the heat density $w$ is proportional to the temperature, i.e., $w = \varphi c_p \theta$, that the heat flux $q$ is proportional to the temperature gradient $\nabla \theta$, i.e., $q = -\kappa \nabla \theta$, and that heat is conserved, i.e., $\partial_t w + \text{div} \ q = 0$. Use the values
\[
\varphi = 1.009 \cdot 10^{-3} \text{ kg/m}^3, \quad \kappa = 0.597 \text{ W/m K}, \quad c_p = 4.186 \cdot 10^3 \text{ J/kg K},
\]
assume that the can is $0.115$ m high, has a diameter of $0.067$ m, and stands upright in the refrigerator with an environmental temperature varying linearly from $4^\circ C$ at the bottom to $5^\circ C$ at the top of the can. Determine the time needed to cool the liquid from $15^\circ C$ below $8^\circ C$ using the Crank–Nicolson method. Discuss the reliability of your result and limitations of the mathematical model.

A.4 Adaptivity

A.4.1 Local Resolution of Corner Singularities

Exercise A.4.1 Show that the triangles in a graded grid of the reference element defined by $J \in \mathbb{N}$ and $\beta \geq 1$ satisfy a minimum angle condition which is independent of $J$. 
Exercise A.4.2 Let \( u \in C^1(\mathbb{R}^2) \) and \( \tilde{u}(r, \phi) = u(r \cos \phi, r \sin \phi) \). Show that we have

\[
\nabla u(x) = \begin{bmatrix} \cos \phi & \sin \phi \\ \sin \phi & -\cos \phi \end{bmatrix} \begin{bmatrix} \partial_r \tilde{u}(r, \phi) \\ r^{-1} \partial_\phi \tilde{u}(r, \phi) \end{bmatrix}
\]

for \( x = r(\cos \phi, \sin \phi) \in \mathbb{R}^2 \setminus \{0\} \). Conclude that \(|\nabla u|^2 = (\partial_r \tilde{u})^2 + r^{-2}(\partial_\phi \tilde{u})^2\).

Exercise A.4.3

(i) For \( x = (x_1, x_2) \in \mathbb{R}^2 \) with \( x_1, x_2 > 0 \), let

\[
r(x_1, x_2) = (x_1^2 + x_2^2)^{1/2}, \quad \phi(x_1, x_2) = \arctan(x_1/x_2).
\]

Show that with the Frobenius norm, we have

\[
|\nabla \phi| \leq \frac{1}{r}, \quad |D^2 \phi| \leq \frac{4}{r^2}, \quad |\nabla r| \leq 1, \quad |D^2 r| \leq \frac{2}{r}.
\]

(ii) In polar coordinates \( (r, \phi) \in \mathbb{R}_{>0} \times (0, \pi/2) \) with respect to the origin in \( \mathbb{R}^2 \) and for \( \alpha > 0 \), let \( u_\alpha(r, \phi) = r^\alpha v(\alpha \phi) \) with a \( 2\pi \)-periodic function \( v \in C^2([0, 2\pi]) \).

For \( x = r(\cos \phi, \sin \phi) \) let \( \tilde{u}_\alpha(x) = u_\alpha(r, \phi) \). Show that there exists \( c_{\alpha, v} > 0 \) such that \(|D^2 \tilde{u}_\alpha(x)| \leq c_{\alpha, v}|x|^{\alpha-2}\).

Exercise A.4.4 For \( \alpha > 0 \) and polar coordinates \( (r, \phi) \) in \( \mathbb{R}^2 \), let \( S(\alpha) = r^\alpha \sin(\alpha \phi) \).

Show that \( S(\alpha) \) is harmonic in the slit domain \( \mathbb{R}^2 \setminus \{(x_1, 0) \in \mathbb{R}^2 : x_1 \geq 0\} \).

Exercise A.4.5 For a graduation strength \( \beta \geq 1 \) and an integer \( J \in \mathbb{N} \), construct a graded grid of the reference element \( T_{\text{ref}} = \text{conv}\{(0,0), (1,0), (0,1)\} \) such that every triangle \( T \) in the triangulation has a right angle, satisfies \( h_T \leq c_\beta/J \), and \( h_T \leq c_\beta J^{-1} |x|^{(\beta-1)/\beta} \) for all \( x \in T \).

Exercise A.4.6 Let \( \omega \subset \mathbb{R}^d \) be open and bounded and \( f \in L^2(\omega) \). Show that for

\[
\bar{f} = |\omega|^{-1} \int_\omega f \, dx
\]

we have \( \|f - \bar{f}\|_{L^2(\omega)}^2 = \min_{c \in \mathbb{R}} \|f - c\|_{L^2(\omega)}^2 = \|f\|_{L^2(\omega)}^2 - |\omega||\bar{f}|^2 \).

Exercise A.4.7 Let \( \Omega \subset \mathbb{R}^2 \) be a bounded, polygonal Lipschitz domain in \( \mathbb{R}^2 \) with corners \( P_1, P_2, \ldots, P_L \in \mathbb{R}^2 \). Devise an algorithm for the generation of a regular family of triangulations that are graded towards re-entrant corners.

Exercise A.4.8 For \( s \in \mathbb{R}_{\geq 0} \) and \( x \in (0,1) \) set \( f(x) = x^s \). Construct a grid \( 0 = x_0 < x_1 < \cdots < x_n = 1 \), such that with the nodal interpolant \( \mathcal{I}_n f \) of \( f \) on the grid we have \( \|f - \mathcal{I}_n f\|_{L^\infty((0,1))} \leq c_n n^{-\frac{1}{2}} \).
Exercise A.4.9 Let $u \in H^1(\Omega)$ be the solution of the Poisson problem $-\Delta u = f$ in $\Omega$ with boundary condition $u|_{\partial\Omega} = u_D$ for a given function $u_D \in C(\partial\Omega)$. For a triangulation $\mathcal{T}_h$ of $\Omega$, let $u_h \in S^1(\mathcal{T}_h)$ be the Galerkin approximation with $u_h(z) = u_D(z)$ for all $z \in \mathcal{N}_h \cap \partial\Omega$. Show that
\[
\|\nabla(u - u_h)\| \leq \|\nabla(u - v_h)\|
\]
for every function $v_h \in S^1(\mathcal{T}_h)$ satisfying $v_h(z) = u_D(z)$ for all $z \in \mathcal{N}_h \cap \partial\Omega$.

Exercise A.4.10 Derive an error estimate for approximating a singularity function $u_{\alpha}(r,\phi) = r^\beta v(\alpha\phi)$ on a graded grid of the reference element with grading strength $\beta > 1/a$.

Quiz A.4.1 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singularity functions $S \in L^1(\Omega)$ associated with a polygonal domain $\Omega \subset \mathbb{R}^2$ satisfy $S \in H^{3/2}(\Omega)$</td>
<td>False</td>
</tr>
<tr>
<td>The number of triangles in a graded grid with grading strength $\beta \geq 1$ and maximal mesh-size $h = 1/J$ is of order $\mathcal{O}(J)$</td>
<td>True</td>
</tr>
<tr>
<td>A graded grid is a quasuniform triangulation with a local resolution towards the origin</td>
<td>True</td>
</tr>
<tr>
<td>The approximation error for a singularity function in $H^1$ is of order $\mathcal{O}(h)$</td>
<td>True</td>
</tr>
<tr>
<td>The approximation error for a corner singularity on a graded grid is up to a constant bounded from below by $h^\beta (1 +</td>
<td>\log(h)</td>
</tr>
</tbody>
</table>

A.4.2 Error Control and Adaptivity

Exercise A.4.11 Let $(\mathcal{T}_h)_{h>0}$ be a sequence of uniformly shape-regular triangulations of the bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$. For each node $z \in \mathcal{N}_h$, let $\omega_z$ be the node patch with diameter $h_z$.

(i) Prove that there exists a constant $c_{loc} > 0$, such that for all $h > 0$, $z \in \mathcal{N}_h$, and $T \in \mathcal{T}_h$ such that $z \in T$, we have
\[
h_T \leq h_z \leq c_{loc} h_T.
\]
Show that this is not true if the condition $z \in T$ is omitted.

(ii) Show that there exists a number $K \geq 0$, such that for all $h > 0$ and $z \in \mathcal{N}_h$ we have
\[
|\{T \in \mathcal{T}_h : z \in T\}| \leq K.
\]
Exercise A.4.12
(i) Let $Q = [0, h]^d \subset \mathbb{R}^d$ and $R = [0, h]^{d-1} \times \{0\}$. Show that there exists $c > 0$ which does not depend on $h > 0$ such that

$$\|v\|_{L^2(R)}^2 \leq c\left(h^{-1}\|v\|_{L^2(Q)}^2 + h\|\nabla v\|_{L^2(Q)}^2\right)$$

for all $v \in H^1(Q)$.

Hint: Use one-dimensional integration-by-parts.

(ii) Let $T \subset \mathbb{R}^d$ be a triangle or tetrahedron and $S \subset \partial T$ be a side of $T$. Show that there exists $c_T > 0$ which does not depend on $h_T = \text{diam}(T)$ such that

$$\|v\|_{L^2(S)}^2 \leq c_T^2\left(h_T^{-1}\|v\|_{L^2(T)}^2 + h_T\|\nabla v\|_{L^2(T)}^2\right)$$

for all $v \in H^1(T)$.

Exercise A.4.13
(i) Show that the Clément quasi-interpolant $\mathcal{J}_h : H^1_0(\Omega) \to \mathcal{S}_1(\mathcal{T}_h)$ is not a projection, i.e., there exists $v_h \in \mathcal{S}_1(\mathcal{T}_h)$ such that $v_h \neq \mathcal{J}_h v_h$.

(ii) Compare approximation results, domains of definition, and projection properties for the nodal interpolant $I_h$ and the Clément quasi-interpolant $\mathcal{J}_h$ related to the finite element space $\mathcal{S}_1(\mathcal{T}_h)$.

Exercise A.4.14 Let $\mathcal{T}_h$ be the triangulation of $\Omega = (0, 1)^2$ consisting of four halved squares with diagonals parallel to the vector $(1, 1)$. Define $u(x, y) = xy$ and $u_h = I_h u$. Compute the jumps $\|\nabla u_h \cdot n_S\|$ for every interior side $S$.

Exercise A.4.15 Let $u_h \in \mathcal{S}_1(\mathcal{T}_h)$ and $S \in \mathcal{T}_h$ such that $S = T_1 \cap T_2$ for $T_1, T_2 \in \mathcal{T}_h$ and let $\tau_S$ be a tangent vector on $S$. Show that

$$\nabla u_h|_{T_1} \cdot \tau_S = \nabla u_h|_{T_2} \cdot \tau.$$

Exercise A.4.16 Let $v \in L^1(\Omega)$ and $\mathcal{J}_h v$ be its Clément interpolant on a regular family of triangulations $(\mathcal{T}_h)_{h > 0}$. Show that $\mathcal{J}_h v \to v$ in $L^1(\Omega)$.

Exercise A.4.17 Derive an a posteriori error estimate for approximating the boundary value problem

$$-\Delta u + u = f \text{ in } \Omega, \quad \partial_n u = g \text{ on } \Gamma_N = \partial \Omega.$$

Exercise A.4.18 Show that there are constants $c_{e,1}, c_{e,2} > 0$ such that for every $h > 0$ and every $T \in \mathcal{T}_h$ with $T = \text{conv}\{z_1, z_2, \ldots, z_{d+1}\}$ the function $b_T = \varphi_{e_1} \varphi_{e_2} \cdots \varphi_{e_{d+1}} \in H^1(\Omega) \cap C(\overline{\Omega})$ satisfies

$$\text{supp } b_T \subset T, \quad \int_T b_T = c_{e,1}|T|, \quad \|\nabla b_T\|_{L^2(T)} \leq c_{e,2}h_T^{d/2 - 1}.$$
Exercise A.4.19

(i) Let $\mathcal{T}_0$ be a triangulation of a Lipschitz domain $\Omega \subset \mathbb{R}^2$ consisting of halved squares only. Prove that every refinement obtained with the red-green-blue refinement strategy leads to a triangulation that consists of right-angled triangles.

(ii) Show that triangulations obtained from an initial triangulation $\mathcal{T}_0$ of $\Omega \subset \mathbb{R}^2$ satisfy a minimum angle condition.

Exercise A.4.20 Let $\mathcal{T}_0$ be the triangulation of $\Omega = (0, 2)^2$ consisting of four halved squares with diagonals parallel to the vector $(1, 1)$. Assume that the triangles containing the points $(2, 1)/3$ and $(5, 4)/3$ are marked for refinement. Determine the refined triangulations with the red-green-blue and the bisection strategy.

Quiz A.4.2 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

| The coefficients $v_z$ that define the Clément quasi-interpolant are obtained as the solution of a linear system of equations |
| The local Poincaré inequality controls the error for approximating a function by its average |
| For finite element functions the Clément interpolant coincides with the nodal interpolant |
| An a posteriori error estimate bounds the approximation error by the jumps of the exact solution and computable terms |
| The efficiency of an error estimator refers to a lower bound for the error up to generic constants |

A.4.3 Convergence of Adaptive Methods

Exercise A.4.21 Let $\mathcal{T}_*$ be a refinement of $\mathcal{T}_h$, i.e., every element in $\mathcal{T}_h$ is the union of elements in $\mathcal{T}_*$. Let $u_h \in H^1_0(\mathcal{T}_h)$ and $u_* \in H^1_0(\mathcal{T}_*)$ be the corresponding Galerkin approximations of the Poisson problem. Prove that $\mathcal{H}^1(\mathcal{T}_h) \subset \mathcal{H}^1(\mathcal{T}_*)$ and

$$\|\nabla (u - u_h)\|^2 = \|\nabla (u - u_*)\|^2 + \|\nabla (u_* - u_h)\|^2.$$ 

Exercise A.4.22 Let $u_h \in H^1_0(\mathcal{T}_h)$ be the Galerkin approximation of the solution $u \in H^1_0(\Omega)$ of the Poisson problem with right-hand side $f \in L^2(\Omega)$. For every $v \in H^1_0(\Omega)$, define

$$\langle \mathcal{R}_{u_h}, v \rangle = \int_\Omega \nabla u_h \cdot \nabla v \, dx - \int_\Omega f v \, dx.$$
Show that for the operator norm
\[ \| R_{u_h} \|_* = \sup_{v \in H^1_0(\Omega) \setminus \{0\}} \frac{\langle R_{u_h}, v \rangle}{\| \nabla v \|} , \]
we have \( \| R_{u_h} \|_* = \| \nabla (u - u_h) \| \).

**Exercise A.4.23** For a triangulation \( T_h \), let \( T_{h/2} \) denote the triangulation obtained from a uniform refinement of \( T_h \). Assume that there exists \( 0 < q < 1 \), such that the corresponding Galerkin approximations of the Poisson problem satisfy
\[ \| \nabla (u - u_{h/2}) \| \leq q \| \nabla (u - u_h) \| . \]
Show that the error estimator
\[ \eta_{h \to h/2}(u_h) = \| \nabla (u_h - u_{h/2}) \| \]
is reliable and efficient. Devise an adaptive algorithm based on this estimator.

**Exercise A.4.24** Construct triangulations \( T_1 \) and \( T_2 \) of \( \mathbb{T}^2 \) and a function \( f \in L^2(\Omega) \), such that the Galerkin approximations \( u_1 \) and \( u_2 \) of the Poisson problem coincide but are different from the true solution. Show that the error does not decrease but the residual error estimator decays in the passage from \( T_1 \) to \( T_2 \).

**Exercise A.4.25** Let \( \varphi_z \in P^1(T_h) \) be the nodal basis function associated with a node \( z \in N_h \). Show that for every \( p \in [1, \infty] \), there exists a constant \( c_z > 0 \) such that
\[ \| \nabla^\ell \varphi_z \|_{L^p(\Omega)} \leq c_z h_z^{d/p-\ell} . \]

**Exercise A.4.26** Let \( z \in N_h \) and \( S_z \in T_h \) be such that \( z \in S_z \).
(i) Show that there exists \( \psi_z \in P^1(S_z) \), such that for all \( z' \in N_h \) we have
\[ \int_{S_z} \psi_z \varphi_{z'} \, ds = \delta_{zz'} \]
(ii) Prove that \( \| \psi_z \|_{L^\infty(S_z)} \leq ch_z^{-(d-1)} \).

**Exercise A.4.27** Let \( \Omega \subset \mathbb{R}^d \) be a bounded Lipschitz domain, and \( T_h \) a triangulation of \( \Omega \). Let \( f \in L^2(\Omega) \) and let \( \tilde{f} \) be the elementwise constant function on \( T_h \) defined by the averages of \( f \) on every element \( T \in T_h \). Let \( u, \tilde{u} \in H^1_0(\Omega) \) be the weak solutions of the Poisson problem with right-hand sides \( f, \tilde{f} \), respectively. Assume that \( f|_T \in H^1(T) \) for all \( T \in T_h \). Show that for a constant \( c_p > 0 \) we have
\[ \| \nabla (u - \tilde{u}) \| \leq c_p h_{\text{max}}^2 \| \nabla \tilde{f} \| . \]
where $h_{\text{max}} > 0$ is the maximal mesh-size in $\mathcal{T}_h$, and $\nabla \varphi f$ the elementwise gradient of $f$.

**Exercise A.4.28** Let $\mathcal{T}_1$, $\mathcal{T}_2$ be triangulations of $\Omega$, such that $\mathcal{T}_2$ is a refinement of $\mathcal{T}_1$. Assume that $\mathcal{T}_1$ and $\mathcal{T}_2$ coincide in the subdomain $\Omega' \subset \Omega$. Show that in general we have $u_1 \neq u_2$ in $\Omega'$ for related Galerkin approximations $u_j \in \mathcal{P}_0^1 (\mathcal{T}_j)$, $j = 1, 2$.

**Exercise A.4.29** Assume that the Poisson problem in $\Omega \subset \mathbb{R}^d$ with homogeneous Dirichlet boundary conditions is $H^2$-regular, i.e., $\|D^2 \psi\| \leq c_2 \|\Delta \psi\|$ for all $\psi \in H^1_0(\Omega) \cap H^2(\Omega)$. Prove that for the approximation error $u - u_h$ of the Poisson problem with the right-hand side $f \in L^2(\Omega)$, we have

$$\|u - u_h\| \leq c \left( \sum_{T \in \mathcal{T}_h} \eta_{2,T}^2(u_h) \right)^{1/2},$$

with the error indicators

$$\eta_{2,T}^2(u_h) = h_T^4 \|f\|_{L^2(T)}^2 + \sum_{S \in \mathcal{T}_h, S \subset \partial T} h_T^3 \|\nabla u_h \cdot n_S\|_{L^2(S)}^2.$$

**Exercise A.4.30** Discuss advantages and disadvantages of finite element methods using uniform and adaptively refined triangulations.

**Quiz A.4.3** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>If $\mathcal{T}_{k+1}$ is a refinement of $\mathcal{T}<em>k$, then we have $|\nabla e</em>{k+1}| &lt; |\nabla e_k|$ for the errors $e_k = u - u_h$</td>
<td>False</td>
</tr>
<tr>
<td>The maximal mesh-size in a sequence of adaptively refined triangulations always tends to zero</td>
<td>True</td>
</tr>
<tr>
<td>The Scott–Zhang quasi-interpolant preserves piecewise affine boundary data</td>
<td>True</td>
</tr>
<tr>
<td>The Scott–Zhang quasi-interpolant is well defined for functions $v \in L^\infty(\Omega)$</td>
<td>True</td>
</tr>
<tr>
<td>For an appropriate marking strategy, the adaptive algorithm always defines a convergent sequence of approximations</td>
<td>True</td>
</tr>
</tbody>
</table>

**A.4.4 Adaptivity for the Heat Equation**

**Exercise A.4.31** Let $X$ be a Banach space and let $f : [a, b] \to X$ be affine. Show that for all $t \in [a, b]$ we have $\|f(t) - f(a)\| \leq \|f(b) - f(a)\|$.
Exercise A.4.32

(i) Show that every weak solution of the heat equation satisfies
\[
\int_0^T \| \partial_t u(t) \|^2 \, dt + \frac{1}{2} \| \nabla u(T) \|^2 = \frac{1}{2} \| \nabla u_0 \|^2.
\]

(ii) Prove that weak solutions for the heat equation are unique.

Exercise A.4.33
For \( \tau > 0 \), a fixed triangulation \( \mathcal{T}_h \) of \( \Omega \), and \( \theta \in [0,1] \), the finite element version of the \( \theta \)-scheme is defined by
\[
(d_t U^j, V) + (\nabla[(1 - \theta)U^{j-1} + \theta U^j], \nabla V) = (f(t_{j+1} + \theta), V)
\]
for \( j = 1, 2, \ldots, J \), and all \( V \in \mathcal{S}_0^1(\mathcal{T}_h) \), where \( t_{j+1} + \theta = (j + 1) \tau \).

(i) Show that the iterates \( (U^j)_{j=0}^J \) are well defined.
(ii) Prove that for every \( \ell = 1, 2, \ldots, J \), we have
\[
\tau \sum_{j=1}^{\ell} \| d_t U^j \|^2 + \frac{1}{2} \| \nabla U^j \|^2 + \frac{\theta - 1}{2} \tau \sum_{j=1}^{\ell} \| \nabla d_t U^j \|^2 \leq \frac{\tau}{2} \sum_{j=1}^{\ell} \| f(t_{j+1} + \theta) \|^2.
\]

Exercise A.4.34
For a triangulation \( \mathcal{T}_h \) of \( \Omega \subset \mathbb{R}^2 \), let \( P_{h,0} : L^2(\Omega) \to \mathcal{S}_0^1(\mathcal{T}_h) \) denote the \( L^2 \)-projection, defined for every \( f \in L^2(\Omega) \) by
\[
(P_{h,0} f, V) = (f, V)
\]
for all \( V \in \mathcal{S}_0^1(\mathcal{T}_h) \).

(i) Prove that \( P_{h,0} \) defines a bounded linear operator \( P_{h,0} : L^2(\Omega) \to L^2(\Omega) \).
(ii) Prove that the discrete Laplacian defines a bijection \( -\Delta_h : \mathcal{S}_0^1(\mathcal{T}_h) \to \mathcal{S}_0^1(\mathcal{T}_h) \).
(iii) Show that the Galerkin approximation \( u_h \in \mathcal{S}_0^1(\mathcal{T}_h) \) of the Poisson problem with the right-hand side \( f \in L^2(\Omega) \) is given by
\[
u_h = (-\Delta_h)^{-1} P_{h,0} f.
\]

Exercise A.4.35
Let \( \mathcal{T}_h \) be a triangulation of \( \Omega \subset \mathbb{R}^2 \) and consider the semidiscretized heat equation
\[
\partial_t U(t) = \Delta_h U(t) - F(t)
\]
for \( t \in [0, T] \) with initial condition \( U(0) = u_0 \) and \( F : [0, T] \to \mathcal{S}_0^1(\mathcal{T}_h) \). Formulate sufficient conditions for the existence and uniqueness of a solution \( U : [0, T] \to \mathcal{S}_0^1(\mathcal{T}_h) \).
Exercise A.4.36 Show that there exists a constant $c > 0$, such that for all $v_h \in \mathcal{S}_0^1(\mathcal{T}_h)$ we have
\[ \| \Delta_h v_h \| \leq c h_{\min}^{-1} \| \nabla v_h \| \]
for all $v_h \in \mathcal{S}_0^1(\mathcal{T}_h)$ with the minimal mesh-size $h_{\min}$ of $\mathcal{T}_h$.

Exercise A.4.37 For a sequence of triangulations $(\mathcal{T}_h)_{h>0}$, let $(u_h)_{h>0}$ be the corresponding sequence of Galerkin approximations $u_h \in \mathcal{S}_0^1(\mathcal{T}_h)$ of the Poisson problem with the right-hand side $f \in L^2(\Omega)$ and exact solution $u \in H_0^1(\Omega)$.

(i) Prove that for every $h > 0$ we have
\[ \| \Delta_h u_h \| \leq \| f \|. \]
(ii) Show that
\[ \Delta_h u_h \to \Delta u \]
in $L^2(\Omega)$ as $h \to 0$.

Exercise A.4.38

(i) Construct a triangulation with at least eight elements for which the marking of one element requires the refinement of all elements in a bisection strategy.
(ii) Construct a triangulation with four interior nodes, for which none of the interior nodes can be coarsened locally, but all nodes together can be removed to obtain a coarsened triangulation, assuming that the vertices opposite the longest edges are newest vertices.

Exercise A.4.39

(i) Bisect the marked elements in the triangulation shown in the left plot of Fig. A.4 using a minimal number of compatible edge patch bisectons.
(ii) Perform as many coarsenings as possible in the triangulation shown in the right plot of Fig. A.4.

Exercise A.4.40 Devise a Crank–Nicolson reconstruction for the case of a right-hand side $f$ that is not piecewise constant but continuously differentiable.

Fig. A.4 Triangulations to be compatibly refined (left) and coarsened (right)
Quiz A.4.4  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weak solutions of the heat equation are not unique in general</td>
<td>True</td>
</tr>
<tr>
<td>The approximation error for the heat equation is bounded by the operator norm of the residual</td>
<td>False</td>
</tr>
<tr>
<td>The difference between a function and its elliptic reconstruction can be controlled with estimates for the Poisson problem</td>
<td>True</td>
</tr>
<tr>
<td>The Crank–Nicolson reconstruction is based on a piecewise quadratic interpolant of the iterates</td>
<td>True</td>
</tr>
<tr>
<td>Mesh coarsening is realized by reversing the most recent refinements of a given triangulation</td>
<td>True</td>
</tr>
</tbody>
</table>

A.4.5  Projects

Project A.4.1  Consider the Poisson problem $-\Delta u = 0$ in the domain

$$\Omega_\gamma = \{ x = r(\cos \phi, \sin \phi) : r > 0, 0 < \phi < \gamma \} \cap (-1, 1)^2$$

with Dirichlet boundary conditions $u = u_D$ on $\Gamma_D = \partial \Omega_\gamma$ so that the exact solution is given by $u(r, \phi) = r^{\gamma/\gamma} \sin(\phi \pi / \gamma)$. Experimentally determine for $\gamma = j\pi / 2$, $j = 1, 2, 3, 4$, an optimal grading strength by computing in each case a sequence of approximation errors $\| \nabla (\mathscr{F}_h u - u_h) \|_{L^2(\Omega)}$.

Project A.4.2  Construct a sequence of triangulations $(\mathcal{T}_h)_{h>0}$ to obtain an optimally convergent sequence of approximations of the Poisson problem with the right-hand side $f = 1$ with homogeneous Neumann and Dirichlet boundary conditions on the corresponding parts of $\partial \Omega$ for the domain $\Omega$ shown in Fig. A.5.

Project A.4.3  Modify the red-green-blue refinement routine so that a new node is created in marked triangles, i.e., marked triangles are refined as indicated in Fig. A.6.

Fig. A.5  Domain $\Omega$ with re-entrant corners and specification of $\Gamma_N$
Project A.4.4 Approximate the Poisson problem $-\Delta u = 1$ in the square $\Omega = (-1, 1)^2$ and in the L-shaped domain $\Omega = (-1, 1)^2 \setminus ((0, 1] \times [-1, 0])$ with homogeneous Dirichlet boundary conditions on $\partial \Omega$ on sequences of uniform triangulations $(T_j)_{j=0,1,...}$. For $j = 0, 1, \ldots$, let $s_j = \|\nabla u_j\|_{L^2(\Omega)}$ and define the extrapolated values $\tilde{s}_j$ for $j \geq 2$ via

$$
\tilde{s}_j = \frac{s_j s_{j-2} - s_{j-1}^2}{s_j - 2s_{j-1} + s_{j-2}}
$$

to obtain an accurate approximation of the unknown value $s = \|\nabla u\|_{L^2(\Omega)}$. Use the obtained value to approximate the errors

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

for sequences of uniformly and adaptively refined triangulations and plot the quantities versus numbers of degrees of freedom with a logarithmic scaling for both axes. Determine the experimental rates of convergence.

Project A.4.5 Consider the Poisson problem $-\Delta u = 1$ in the L-shaped domain $\Omega = (-1, 1)^2 \setminus ((0, 1] \times [-1, 0])$ with homogeneous Dirichlet boundary conditions on $\partial \Omega$. For a sequence of uniform triangulations, compute upper bounds for the approximation errors and determine the CPU-time needed to solve each problem. Approximate the problems adaptively, compute error bounds, and compare the CPU-times needed to obtain comparable error bounds. Discuss the benefits of adaptivity on the basis of your results.

Project A.4.6 Let $T_h$ be a triangulation of $\Omega$ and let $T_{h/2}$ be the triangulation obtained from $T_h$ by a uniform red refinement. The $h-h/2$-estimator is defined by

$$
\eta_{h-h/2}(T_h) = \|\nabla (u_h - u_{h/2})\|_{L^2(\Omega)},
$$

where $u_h$ and $u_{h/2}$ are the finite element approximations corresponding to the triangulations $T_h$ and $T_{h/2}$, respectively. Test and compare the error estimator to the residual error estimator for a Poisson problem in a nonconvex domain.

Project A.4.7 Let $u_h \in \mathcal{S}(T_h)$ be a finite element function and define $\omega_h [\nabla u_h] = p_h \in \mathcal{S}(T_h)^d$ by

$$
p_h(z) = |\omega_z|^{-1} \int_{\omega_z} \nabla u_h \, dx
$$
for all nodes \( z \in \mathcal{N}_h \) with node patches \( \omega_z = \text{supp} \varphi_z \). The averaging estimator is defined by

\[
\eta_{\mathcal{A}}(\mathcal{T}_h) = \left\| \nabla u_h - \mathcal{A}_h[\nabla u_h] \right\|_{L^2(\Omega)}.
\]

Test and compare the error estimator to the residual error estimator for a Poisson problem in a nonconvex domain.

**Project A.4.8** Consider the Poisson problem \(-\Delta u = f\) in \( \Omega = (0, 1)^2 \), \( u|_{\partial \Omega} = 0 \), with the function \( f \in L^2(\Omega) \) defined by \( f = -\Delta u \) for

\[
u(x, y) = x(1 - x)y(1 - y) \arctan \left( 50(x, y) - 1 \right),
\]

\[
r(x, y) = \left( (x - 5/4)^2 + (y + 1/4)^2 \right)^{1/2}.
\]

Determine errors and error estimates on sequences of uniformly and adaptively refined triangulations and plot them versus degrees of freedom in a double-logarithmic scaling. Comment on possible benefits of adaptivity.

## A.5 Iterative Solution Methods

### A.5.1 Multigrid

**Exercise A.5.1** Let \( A_h \) be the stiffness matrix related to a finite element space \( \mathcal{X}_0^1(\mathcal{T}_h) \). Show that the estimate \( \text{cond}_2(A_h) \leq c h^{-2} \) is optimal in the sense that there exists a constant \( c' > 0 \) such that \( \text{cond}_2(A_h) \geq c' h^{-2} \). Consider the case \( d = 1 \) first.

**Exercise A.5.2** Let \( V \) be a finite-dimensional space, let \( \langle \cdot, \cdot \rangle_V : V \times V \to \mathbb{R} \) be a scalar product on \( V \), and let \( a : V \times V \to \mathbb{R} \) be a symmetric bilinear form on \( V \). Show that there exists an orthonormal basis \( (v_1, v_2, \ldots, v_n) \) of \( V \), and numbers \( \lambda_1, \lambda_2, \ldots, \lambda_n \) such that

\[
 a(v_j, w) = \lambda_j \langle v_j, w \rangle_V
\]

for all \( w \in V \).

**Exercise A.5.3** Construct an inner product \( \langle \cdot, \cdot \rangle \) on \( \mathbb{R}^n \) and a symmetric matrix \( A \in \mathbb{R}^{n \times n} \) such that \( A \) is not symmetric with respect to this inner product, i.e., there exist \( V, W \in \mathbb{R}^n \) such that

\[
 \langle AV, W \rangle \neq \langle V, AW \rangle.
\]
Exercise A.5.4 Prove that for every \( k \in \mathbb{N} \) we have
\[
\max_{t \in [0,1]} t(1-t)^k \leq \frac{1}{ek} \leq k^{-1}.
\]

Exercise A.5.5 Let \( A_h \) be the system matrix resulting from a \( P_1 \)-finite element discretization of the elliptic boundary value problem

\[-\Delta u + \alpha u = f \text{ in } \Omega, \quad u|_{\partial \Omega} = 0.\]

Derive an upper bound for the condition number with a precise dependence on the parameter \( \alpha \geq 0 \).

Exercise A.5.6 Let \( A_h \in \mathbb{R}^{n \times n} \) be the finite element stiffness matrix related to the Poisson problem, i.e., for \( v_h, w_h \in \mathcal{S}_1^0(\mathcal{T}_h) \) with coefficient vectors \( V_h, W_h \in \mathbb{R}^n \), we have

\[V_h^T A_h W = \int_\Omega \nabla v_h \cdot \nabla w_h \, dx.\]

For a diagonalization \( A_h = Q^T D Q \) with a diagonal matrix \( D \in \mathbb{R}^{n \times n} \) and an orthogonal matrix \( Q \in \mathbb{R}^{n \times n} \), and \( s \in \mathbb{R} \), we define

\[A^s_h = Q^T D^s Q.\]

(i) Show that the expression \( \|v_h\|_s = \|A^s_h^{-1/2} V_h\| \) defines a norm on \( \mathcal{S}_1^0(\mathcal{T}_h) \) which coincides with the Euclidean norm if \( s = 0 \) and with the \( H^1 \) norm if \( s = 1 \).

(ii) Show that for \( s, r, t \) with \( s = (r + t)/2 \), we have \( \|v_h\|_s \leq \|v_h\|^{1/2}_t \|v_h\|^{1/2}_r \).

(iii) Show that if \( A_h U_h = B_h \), then we have \( \|u_h\|_{s+2} = \|b_h\|_s \).

Exercise A.5.7 Let \( \mathcal{T}_h \) be a triangulation of \( \Omega \), and let \( A_h \in \mathbb{R}^{n \times n} \) be the stiffness matrix defined by the space \( \mathcal{S}_1^0(\mathcal{T}_h) \). Construct norms \( \| \cdot \|_s \) and \( \| \cdot \|_r \) on \( \mathbb{R}^n \) such that for the induced condition number we have \( \text{cond}_{\ell_1}(A_h) = 1 \).

Exercise A.5.8 Let \( \mathcal{T}_h \) be a triangulation that is obtained from a triangulation \( \mathcal{T}_H \) by a red-green-blue-refinement procedure. Show that there exists a uniquely defined prolongation operator \( P : \mathbb{R}^N \to \mathbb{R}^n \), such that \( PV_H \) is the coefficient vector of the function \( v_H \in \mathcal{S}_1^0(\mathcal{T}_H) \) with respect to the nodal basis in \( \mathcal{S}_1^0(\mathcal{T}_h) \), provided that \( v_H \) is defined by the coefficient vector \( V_H \in \mathbb{R}^N \).

Exercise A.5.9 Let \( P : \mathcal{S}_1^0(\mathcal{T}_H) \to \mathcal{S}_1^0(\mathcal{T}_h) \) be the prolongation operator between two nested finite element spaces. Show that the transpose \( P^T \) defines an operator \( \mathcal{S}_1^0(\mathcal{T}_h) \to \mathcal{S}_1^0(\mathcal{T}_H) \) that is not the inverse of \( P \) and does not coincide with the nodal interpolation on \( \mathcal{T}_H \).
Exercise A.5.10 Assume that for norms \( \| \cdot \|_X \) and \( \| \cdot \|_Y \) we have the smoothing property

\[
\| \mathcal{S}^k v_h \|_X \leq c_1 h^{-\beta} k^{-\gamma} \| v_h \|_Y,
\]

for an iterative method for solving an equation \( A_h u_h = b_h \), and the approximation property

\[
\| v_h - v_{2h} \|_Y \leq c_2 h^\beta \| v_h \|_X,
\]

for all \( v_h \in V_h \), and an appropriate \( v_{2h} \in V_{2h} \). Devise and analyze an abstract two-level method for the efficient numerical solution of the equation \( A_h u_h = b_h \). Discuss the computational complexity, and apply the framework to the approximation of the Poisson problem.

Quiz A.5.1 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>The condition number of a finite element stiffness matrix depends on the choice of the basis</td>
<td>True/False</td>
</tr>
<tr>
<td>To achieve a reduction to 5% of the initial error, approximately 20 Richardson iterations are necessary</td>
<td>True/False</td>
</tr>
<tr>
<td>The smallest eigenvalue of the discrete Laplace operator is of order ( O(h^d) )</td>
<td>True/False</td>
</tr>
<tr>
<td>The prolongation operator computes the coefficients of a coarse grid function on a finer grid</td>
<td>True/False</td>
</tr>
<tr>
<td>The coarse grid correction ( c_H ) is the Galerkin approximation of the Poisson problem (-\Delta u = f ) in ( \Omega ), ( u</td>
<td>_{\partial\Omega} = 0 )</td>
</tr>
</tbody>
</table>

A.5.2 Domain Decomposition

Exercise A.5.11 Let \( f \in L^2(\Omega) \) and \( u \in H^2(\Omega) \cap H^1_0(\Omega) \). Show that we have

\[-\Delta u = f \text{ in } \Omega, \quad u|_{\partial\Omega} = 0,\]

if and only if the functions \( u_j = u|_{\Omega_j}, j = 1, 2, \) satisfy

\[-\Delta u_j = f \text{ in } \Omega_j, \quad u_j|_{\Gamma_j} = 0,\]

for \( j = 1, 2 \) and

\[ u_1 = u_2, \quad \partial_{n_1} u_1 = -\partial_{n_2} u_2. \]
on $\gamma$, where $\partial_j u_j = \nabla u_j \cdot n_j$ on $\gamma$ with the outer unit normal $n_j$ to $\partial \Omega_j$, i.e., $n_2 = -n_1$.

**Exercise A.5.12** Show that if $\Omega = (0, 1), f = 0$, and $\Omega_1 = (0, a), \Omega_2 = (a, 1)$ for $0 < a < 1/2$, then the Dirichlet–Neumann method converges if and only if $\theta < 1$.

**Exercise A.5.13** Prove that every stationary pair $(u_1, u_2)$ for the Dirichlet–Neumann method coincides with the solution $u \in H^1_0(\Omega)$ of the Poisson problem with the right-hand side $f$, i.e., $u_j = u|_{\Omega_j}$ for $j = 1, 2$.

**Exercise A.5.14** Generalize the Dirichlet–Neumann method to partitions with more than two subdomains.

**Exercise A.5.15** Let $v \in H^1(\Omega)$ and $\partial \Omega = \Gamma \cup \gamma$, for disjoint sets $\Gamma$ and $\gamma$, and assume that $v|_\gamma = 0$. Show that if $w \in H^1(\Omega)$ we have

$$-\Delta w = 0 \text{ in } \Omega, \quad w|_\Gamma = 0, \quad w|_\gamma = v|_\gamma$$

if and only if $\|\nabla w\|_{L^2(\Omega)} \leq \|\nabla (w + \phi)\|_{L^2(\Omega)}$ for all $\phi \in H^1(\Omega)$ with $\phi|_\gamma = 0$.

**Exercise A.5.16** Let $\Omega_1, \Omega_2$ be a nonoverlapping partition of $\Omega$ with interface $\gamma = \partial \Omega_1 \cap \partial \Omega_2$.

(i) Construct diffeomorphisms $\Phi_j : \Omega_j \to \omega_j$ for $j = 1, 2$, with sets $\omega_1 \subset \Omega_2$ and $\omega_2 \subset \Omega_1$ such that $\Phi_j(\gamma) = \gamma$ for $j = 1, 2$.

(ii) Show that the expressions $\|\psi\|_j = \|\nabla H_j \psi\|_{L^2(\Omega_j)}$ with the harmonic extension $H_j \psi$ of $\psi$ to $\Omega_j$, $j = 1, 2$, define equivalent norms on $H^{1/2}_{00}(\gamma)$.

**Exercise A.5.17** Let $\Omega_1, \Omega_2$ be a nonoverlapping partition of $\Omega$ with interface $\gamma = \partial \Omega_1 \cap \partial \Omega_2$, and set $\Gamma_j = \partial \Omega_j \cap \partial \Omega$. Define $T : H^{1/2}_{00}(\gamma) \to H^{1/2}_{00}(\gamma)$ by $T \psi = w_2|_\gamma$, where $w_2 \in H^1_{\Gamma_2}(\Omega_2)$ solves

$$-\Delta w_2 = 0 \text{ in } \Omega_2, \quad w_2|_{\Gamma_2} = 0, \quad \partial_n w_2 = -\partial_n H_1 \psi.$$ 

Show that this is equivalent to

$$a_2(w_2, v_2) = -a_1(H_1 \psi, H_1 v_2|_\gamma)$$

for all $v_2 \in H^1_{\Gamma_2}(\Omega_2)$.

**Exercise A.5.18** Compute the iterates $(u^k_1, u^k_2)_{k=0,\ldots,5}$ of the alternating Schwarz method for the problem $-u'' = 1, u(0) = u(1) = 0$ for $\Omega_1 = (0, 1/2 + \delta)$ and $\Omega_2 = (1/2 - \delta, 1)$ for $0 < \delta < 1/2$.

**Exercise A.5.19** Prove that with the appropriate spaces $V_1, V_2 \subset V$, the alternating Schwarz method can be equivalently defined by the problems of computing $w^k_1 \in V_1$ such that

$$a(w^k_1, v_1) = b(v_1) - a(u^k, v_1)$$
for all $v_1 \in V_1$, setting $u^{k+1/2} = u_k + w_1^k$, computing $w_2^k \in V_2$ such that

$$a(w_2^k, v_2) = b(v_2) - a(u^{k+1/2}, v_2)$$

for all $v_2 \in V_2$, and setting $u^{k+1} = u^{k+1/2} + w_2^k$.

**Exercise A.5.20** For a Hilbert space $V$ and a subspace $W \subset V$, let $\mathcal{P} : V \to V$ denote the orthogonal projection onto $W$. Prove that $\mathcal{P} \circ (1 - \mathcal{P}) = 0$, and that the images of $\mathcal{P}$ and $(1 - \mathcal{P})$ are orthogonal subsets of $V$.

**Quiz A.5.2** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

If $u \in H^1(\Omega)$ and $\Omega = \overline{\Omega}_1 \cup \overline{\Omega}_2$, $\gamma = \partial \Omega_1 \cap \partial \Omega_2$, then for $u_j = u|_{\Omega_j}$, $j = 1, 2$, we have $u_1|_{\gamma} = u_2|_{\gamma}$

If $u \in H^1(\Omega_1 \cup \Omega_2)$, $\gamma = \partial \Omega_1 \cap \partial \Omega_2$, then for $u_j = u|_{\Omega_j}$, $j = 1, 2$, we have $u_1|_{\gamma} = u_2|_{\gamma}$

if $v \in H^1(\Omega)$, $\gamma \subset \partial \Omega$, then the harmonic extension $w \in H^1(\Omega)$ of $v|_{\gamma}$ to $\Omega$ with $w|_{\partial \Omega \setminus \gamma} = 0$ is well defined

The overlapping Schwarz method converges only if the subdomains are nonfloating

The Poisson problem in $\Omega$ can be decomposed into independent problems on subdomains

### A.5.3 Preconditioning

**Exercise A.5.21** Let $A \in \mathbb{R}^{n \times n}$ be symmetric and positive definite and $b \in \mathbb{R}^n$. Assume that $C \in \mathbb{R}^{n \times n}$ and $K \in \mathbb{R}^{n \times n}$ are regular matrices with $C = KK^T$. Apply the conjugate gradient algorithm to the linear systems of equations $(CA)x =Cb$ and $(K^TAK)K^{-1}x = K^Tb$ and show that the resulting methods are equivalent to the preconditioned CG algorithm. Compare the cost of the matrix vector products $z \mapsto Cz$ and $z \mapsto (KK^T)z$.

**Exercise A.5.22** Let $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ be a scalar product on $\mathbb{R}^n$. Show that the minimal and maximal eigenvalues of a symmetric matrix $A \in \mathbb{R}^{n \times n}$ are given as the extrema of the function

$$R : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}, \quad x \mapsto \frac{\langle Ax, x \rangle}{\langle x, x \rangle}.$$
Exercise A.5.23  Let $A, C \in \mathbb{R}^n$ be symmetric regular matrices. Prove that the extremal eigenvalues of $CA$ are the extremal of the mapping

$$R : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}, \quad x \mapsto \frac{Ax \cdot x}{C^{-1}x \cdot x}.$$

Exercise A.5.24  Let $\mathcal{T}_h$ be a triangulation of $\Omega \subset \mathbb{R}^d$ with nodes $(z_1, z_2, \ldots, z_n) = \mathcal{N}_h$. Let $M$ and $\tilde{M}$ be the matrices in $\mathbb{R}^{n \times n}$ defined by

$$M_{ij} = \int_{\Omega} \varphi_{z_i} \varphi_{z_j} \, dx, \quad \tilde{M}_{ij} = \int_{\Omega} \mathcal{I}_h[\varphi_{z_i} \varphi_{z_j}] \, dx,$$

for $i,j = 1, 2, \ldots, n$, with the nodal interpolation operator $\mathcal{I}_h : C(\overline{\Omega}) \to \mathcal{P}^1(\mathcal{T}_h)$.

(i) Show that $M$ and $\tilde{M}$ are positive definite and symmetric, and that $\tilde{M}$ is diagonal.

(ii) Prove that for every $v \in \mathbb{R}^n$ we have

$$v^T M v \leq v^T \tilde{M} v \leq (d + 2) v^T M v.$$

*Hint:* Prove the inequality on every element $T \in \mathcal{T}_h$ first.

(iii) Show that $C = \tilde{M}^{-1}$ is an optimal preconditioner for $M$, i.e.,

$$\text{cond}_2(CM) \leq d + 2,$$

and the evaluation $r \mapsto Cr$ is of complexity $O(n)$.

Exercise A.5.25  Let $0 < \gamma < 1$ and define $\Gamma_{ij} = \gamma^{i-j}/2$ for $i,j = 0, 1, \ldots, L$.

(i) Prove that $\varrho(\Gamma) \leq 1/(1 - \gamma^{1/2})$.

(ii) Show that for vectors $\alpha, \beta \in \mathbb{R}^{L+1}$ we have

$$\gamma^{i-j}/2 \alpha_i \beta_j \leq \varrho(\Gamma) |\alpha||\beta|.$$

Exercise A.5.26  Let $A \in \mathbb{R}^{n \times n}$ be regular and define the diagonal matrix $D \in \mathbb{R}^{n \times n}$ by $d_{ii} = \sum_{j=1}^{n} |a_{ij}|$ for $i = 1, 2, \ldots, n$. Show that for every diagonal matrix $T \in \mathbb{R}^n$, we have

$$\text{cond}_{\infty}(D^{-1}A) \leq \text{cond}_{\infty}(TA),$$

where $\text{cond}_{\infty}(A) = \|A\|_{\infty} \|A^{-1}\|_{\infty}$ is the condition number defined by the row sum norm $\|A\|_{\infty} = \max_{i=1,\ldots,n} \sum_{j=1}^{n} |a_{ij}|$.

Exercise A.5.27  Derive an upper bound for the computational complexity of the application of the BPX preconditioner, using that the number of levels satisfies $L \sim \log_2(h^{-1})$. 
Exercise A.5.28 Let \( V_0 \subset V_1 \subset \cdots \subset V_L = V \subset L^2(\Omega) \). Let \( v \in V \) and construct functions \( w_0, w_1, \ldots, w_L \in V \) with \( w_\ell \in V_\ell \) for \( \ell = 0, 1, \ldots, L \), such that \( w_{\ell+1} = w_{\ell+2} = \cdots = w_L = 0 \) of \( v \in V_\ell \), and
\[
\sum_{\ell=0}^L w_\ell = v, \quad \left\| \sum_{\ell=0}^L w_\ell \right\|_{L^2(\Omega)}^2 = \sum_{\ell=0}^L \left\| w_\ell \right\|_{L^2(\Omega)}^2.
\]

Exercise A.5.29 Show that if the assumption about the \( H^2 \)-regularity of the Poisson problem is omitted, then for the BPX preconditioner \( C \in \mathbb{R}^{n \times n} \) we have that
\[
\text{cond}_2(CA) \leq c \log_2(h^{-1}).
\]

Exercise A.5.30 Show that \( q : \mathbb{R}^n \to \mathbb{R} \) is a quadratic form, i.e., \( q(\lambda v) = \lambda^2 q(v) \) and \( q(v + w) + q(v - w) = 2q(v) + 2q(w) \) for all \( v, w \in \mathbb{R}^n \) and \( \lambda \in \mathbb{R} \), if and only if there exists a symmetric matrix \( M \in \mathbb{R}^{n \times n} \) such that \( q(v) = v^T M v \) for all \( v \in \mathbb{R}^n \).

Quiz A.5.3 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>The Gauss–Seidel preconditioner is optimal for diagonal matrices</th>
</tr>
</thead>
<tbody>
<tr>
<td>The BPX preconditioner is optimal on regular sequences of triangulations if the Poisson problem is ( H^2 )-regular</td>
</tr>
<tr>
<td>If a preconditioner is optimal, i.e., ( \text{cond}_2(CA) \leq c ), then the preconditioned CG algorithm terminates within a finite number of iterations, independently of the problem size</td>
</tr>
<tr>
<td>If the eigenvalues of the symmetric and positive definite matrices ( A ) and ( C ) satisfy ( c_1 \lambda_i \leq \mu_i^{-1} \leq c_2 \lambda_i ) for ( i = 1, 2, \ldots, n ), then ( \text{cond}_2(CA) \leq c_1/c_2 )</td>
</tr>
<tr>
<td>The strengthened Cauchy–Schwarz inequality states that functions ( v_k, \ell ) in nested spaces ( \mathcal{S}_0(\mathcal{T}_h) \subset \mathcal{S}_0(\mathcal{T}) ) are nearly orthogonal</td>
</tr>
</tbody>
</table>

A.5.4 Projects

Project A.5.1 Let \( \mathcal{T}_h \) be a triangulation of \( \Omega = (0, 1)^2 \). Use the von Mises power method to determine nontrivial functions \( u_h^{(1)}, u_h^{(N)} \in \mathcal{S}_0(\mathcal{T}_h) \) such that there exist \( \lambda_1, \lambda_N \in \mathbb{R} \) with \( \lambda_1 < \lambda_N \) such that
\[
\int_{\Omega} \nabla u_h^{(i)} \cdot \nabla v_h \, dx = \lambda_i \int_{\Omega} u_h^{(i)} v_h \, dx
\]
for all $v_h \in \mathcal{S}_0^1(\mathcal{T}_h)$ and $i = 1, N$. Plot the functions for different meshes and plot the ratios $q_h = \lambda_N / \lambda_1$ for a sequence of refined triangulations versus the mesh-size $h$ using a logarithmic scaling, and determine a relation $q_h \sim h^a$.

**Project A.5.2** Use a sequence of uniformly refined triangulations $(\mathcal{T}_j)_{j=0,1,...}$ and approximately solve the Poisson problem $-\Delta u = 1$, $u|_{\partial \Omega} = 0$, for $\Omega = (0, 1)^d$, $d = 2, 3$, using the backslash operator, a Gauss–Seidel iteration, and the multigrid algorithm. Compare the CPU times and discuss the results.

**Project A.5.3** The W-cycle of the multigrid method is defined by applying the multigrid function twice on every level. Explain the name of the method and experimentally compare its performance with the V-cycle. Experimentally determine a sufficient number of post-smoothing steps by comparing the multigrid approximations with the discrete solution obtained with the backslash operator.

**Project A.5.4** Consider the Poisson problem $-\Delta u = 1$, $u|_{\partial \Omega} = 0$, in the domain $\Omega = (0, 1)^d$, $d = 2, 3$. Use a sequence of uniformly refined triangulations and solve the resulting linear systems of equations of the finite element approximations using the preconditioned conjugate gradient algorithm with Jacobi, equilibration, symmetric Gauss–Seidel, incomplete Cholesky, and BPX preconditioner. Repeat the experiment for a two-dimensional L-shaped domain and discuss your results.

**Project A.5.5** Implement the two-level preconditioner and test it for a Poisson problem and a sequence of uniformly refined triangulations.

**Project A.5.6** Implement the overlapping Schwarz domain decomposition method, test it for two Poisson problems with different domains $\Omega \subset \mathbb{R}^2$, and determine experimentally the dependence of the number of iterations on the geometry of the overlap region.

**Project A.5.7** Generalize the Dirichlet–Neumann method to the case of arbitrarily many nonfloating subdomains and experimentally investigate its convergence for two Poisson problems in $\mathbb{R}^2$ with at least four subdomains.

**Project A.5.8** Let $(\Omega_j)_{j=1,2}$ be a nonoverlapping partition of $\Omega \subset \mathbb{R}^2$ with interface $\gamma$. For a triangulation $\mathcal{T}_h$ of $\Omega$ such that $\Omega_1$ and $\Omega_2$ are matched by unions of triangles in $\mathcal{T}_h$, let $\mathcal{T}_h^{(j)}$ be the induced triangulations of $\Omega_j$, $j = 1, 2$. Let $W_h = \mathcal{S}_0^1(\mathcal{T}_h)|_\gamma$ be the discrete trace space on the interface and defined norms $\| \cdot \|_{j,h}$ for $j = 1, 2$ for $\psi_h \in W_h$ via

$$\| \psi_h \|_{j,h} = \| \nabla u_h^{(j)} \|_{L^2(\Omega_j)},$$

where $u_h^{(j)} \in \mathcal{S}^1(\mathcal{T}_h^{(j)})$ satisfies $u_h^{(j)} = \psi_h$ on $\gamma$, $u_h^{(j)}|_{\partial \Omega_j \setminus \gamma} = 0$, and

$$\int_{\Omega_j} \nabla u_h^{(j)} \cdot \nabla v_h \, dx = 0$$
for all $v_h \in \mathcal{V}_0^1(\mathcal{T}_h^{(j)})$. Define an eigenvalue problem to determine mesh-dependent constants $\lambda_1^h, \lambda_N^h$ such that

$$\lambda_1^h \|\psi_h\|_{1,h} \leq \|\psi_h\|_{2,h} \leq \lambda_N^h \|\psi_h\|_{1,h}$$

for all $\psi_h \in \mathcal{V}_0^1(\mathcal{T}_h)$. Compute the numbers $\lambda_1^h, \lambda_N^h$ for three different partitions of $\Omega = (0, 1)^2$ and three different triangulations in each case.

### A.6 Saddle-Point Problems

#### A.6.1 Discrete Saddle-Point Problems

**Exercise A.6.1** Let $M \in \mathbb{R}^{n \times n}$ and $\| \cdot \|_\ell$ be a norm on $\mathbb{R}^n$ with dual norm $\| \cdot \|_{\ell'}$. Show that $k \geq 0$ is the smallest constant such that

$$x^T My \leq k \|x\|_\ell \|y\|_{\ell'}$$

if and only if

$$k = \sup_{x \in \mathbb{R}^n} \frac{\|Mx\|_{\ell'}}{\|x\|_\ell}.$$

**Exercise A.6.2**

(i) Show that $M \in \mathbb{R}^{n \times n}$ is regular if and only if

$$\inf_{v \in \mathbb{R}^n \setminus \{0\}} \sup_{w \in \mathbb{R}^n \setminus \{0\}} \frac{v^T M w}{\|v\| \|w\|} > 0.$$

(ii) Show that $B \in \mathbb{R}^{m \times n}$ is surjective if and only if

$$\inf_{q \in \mathbb{R}^m \setminus \{0\}} \sup_{v \in \mathbb{R}^n \setminus \{0\}} \frac{q^T B v}{\|q\| \|v\|} > 0.$$

**Exercise A.6.3** Let $f : V \rightarrow W$ be a linear mapping between the $n$-dimensional linear space $V$ and the $n$-dimensional linear space $W$. Show that there exist bases $(v_1, v_2, \ldots, v_n)$ for $V$ and $(w_1, w_2, \ldots, w_m)$ for $W$, such that the matrix $B \in \mathbb{R}^{m \times n}$ representing $f$ in these bases satisfies

$$\ker B = \{v = (v_1, v_2, \ldots, v_n) \in \mathbb{R}^n : v_1 = v_2 = \cdots = v_r = 0\},$$

where $r = \dim \ker B$. 
Exercise A.6.4 Systematically investigate the regularity of the saddle-point matrix

\[ M = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 2 \\ 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 3 & 1 \\ 0 & 0 & 3 & 9 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \]

Exercise A.6.5 Let \( A \in \mathbb{R}^{n \times n} \) be positive semidefinite and symmetric, and let \( K \subset \mathbb{R}^n \) be a linear subspace. Show that

\[ \inf_{v \in K \setminus \{0\}} \sup_{w \in K \setminus \{0\}} \frac{v^TAw}{\|v\|\|w\|} \geq \alpha \]

if and only if \( v^TAv \geq \alpha \|v\|^2 \) for all \( v \in K \).

Exercise A.6.6

(i) For \( B \in \mathbb{R}^{n_B \times n_A} \) let \( B_I \) denote the restriction of \( B \) to \((\ker B)^\perp\). Prove that invertibility of \( B_I : \text{Im} B^T \leftrightarrow \mathbb{R}^{n_B} \) is equivalent to the implication

\[ B^Tz = 0 \quad \implies \quad z = 0, \]

(ii) Show that bijectivity of \( A \in \mathbb{R}^{n_A \times n_A} \) on \( \ker B \) for a matrix \( B \in \mathbb{R}^{n_B \times n_A} \) is equivalent to the implication

\[ y^TAv = 0 \quad \text{for all} \quad v \in \ker B \quad \implies \quad y = 0. \]

Exercise A.6.7 Let \( \| \cdot \|_V \) and \( \| \cdot \|_Q \) be norms on \( \mathbb{R}^{n_A} \) and \( \mathbb{R}^{n_B} \) with duals \( \| \cdot \|_{V'} \) and \( \| \cdot \|_{Q'} \), respectively.

(i) Show that the mapping

\[ \| \cdot \|_\varepsilon : \mathbb{R}^{n_A} \times \mathbb{R}^{n_B} \to \mathbb{R}, \quad (y, z) \mapsto \|y\|_V + \|z\|_Q, \]

defines a norm on \( \mathbb{R}^{n_A} \times \mathbb{R}^{n_B} \), its dual norm is given by

\[ \|(f, g)\|_{\varepsilon'} = \max \{\|f\|_{V'}, \|g\|_{Q'}\} \]

for all \( (f, g) \in \mathbb{R}^{n_A} \times \mathbb{R}^{n_B} \), and that

\[ \|(f, g)\|_{\varepsilon'} \leq \|f\|_{V'} + \|g\|_{Q'} \leq 2\|(f, g)\|_{\varepsilon'}. \]
(ii) Let \( A \in \mathbb{R}^{n_A \times n_A} \) and \( B \in \mathbb{R}^{n_B \times n_A} \) be such that there exist constants \( k_A, k_B \geq 0 \) so that
\[
v^\top A y \leq k_A \|v\|_V \|y\|_V, \quad v^\top B z \leq k_B \|v\|_V \|z\|_Q
\]
for all \( v, y \in \mathbb{R}^{n_A} \) and \( z \in \mathbb{R}^{n_B} \). Show that for the associated saddle-point matrix \( M \) we have with \( \| \cdot \|_r = \| \cdot \|_{\ell r} \) that
\[
\|M\|_{\ell r} \leq k_A + 2k_B.
\]

**Exercise A.6.8** Let \( \| \cdot \|_V \) and \( \| \cdot \|_Q \) be norms on \( \mathbb{R}^{n_A} \) and \( \mathbb{R}^{n_B} \) with dual norms \( \| \cdot \|_{V'} \) and \( \| \cdot \|_{Q'} \), respectively, and let the restriction \( B_I \in \mathbb{R}^{n_B \times n_B} \) of \( B \) to \( (\ker B)^\perp \) be regular.

(i) Show that for every \( z \in \mathbb{R}^{n_B} \) we have
\[
\|z\|_Q = \sup_{s \in \mathbb{R}^{n_A}} \frac{s^\top z}{\|s\|_{Q'}}.
\]

(ii) Show that we have
\[
\|B_I^{-1}\|_{Q'} = \|(B_I^{-1})^{-1}\|_{V'Q'}.
\]

**Exercise A.6.9** Assume that \( a : V \times V \to \mathbb{R} \) is a bounded and coercive bilinear form with constants \( k_a, \alpha > 0 \), \( V_h = \text{span}\{v_1, v_2, \ldots, v_n\} \) a finite-dimensional subspace, and \( M_{jk} = a(v_j, v_k) \) for \( j, k = 1, 2, \ldots, n \).

(i) Show that a norm \( \| \cdot \|_{\ell} \) on \( \mathbb{R}^n \) is defined by
\[
\|x\|_\ell = \left\| \sum_{i=1}^n x_i v_i \right\|_V.
\]

(ii) Show that we have
\[
\frac{y^\top M x}{\|x\|_\ell \|y\|_\ell} = \frac{a(u_h, v_k)}{\|u_h\|_V \|v_k\|_V}.
\]

(iii) Show that for \( \| \cdot \|_r = \| \cdot \|_{\ell r} \) we have \( \text{cond}_r(M) \leq k_a/\alpha \).

**Exercise A.6.10** Let \( A \in \mathbb{R}^{n_A \times n_A} \) be symmetric and positive definite and \( B \in \mathbb{R}^{n_B \times n_A} \). Prove that the Schur complement \( S = BA^{-1}B^\top \in \mathbb{R}^{n_B \times n_B} \) is symmetric and positive definite.
A.6 Saddle-Point Problems

Quiz A.6.1 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>A necessary condition for the unique solvability of a saddle-point problem is that ( n_D \leq n_A ) and ( B ) has maximal rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>If ( \bar{B} ) is injective, then for all ( f \in (\ker B)^\perp ) there exists a unique ( z ) with ( \bar{B}^T z = f )</td>
</tr>
<tr>
<td>The inf-sup condition bounds the operator norm of the left-inverse of a matrix</td>
</tr>
<tr>
<td>If ( B ) is surjective, then it has a left inverse ( B^{-\ell} )</td>
</tr>
<tr>
<td>The symmetric matrix ( A ) defines a bijection on the subspace ( K ) if and only if ( A ) is positive definite on ( K )</td>
</tr>
</tbody>
</table>

A.6.2 Continuous Saddle-Point Problems

Exercise A.6.11

(i) For \( n \in \mathbb{N} \) let \( L_n : \mathbb{R}^n \to \mathbb{R}^n \) be defined by \( (x_1,x_2,\ldots,x_n) \mapsto (x_1,x_2/2,x_3/3,\ldots,x_n/n) \). Determine the condition number of \( L_n \) with respect to the Euclidean norm.

(ii) Let \( \ell^2(\mathbb{N}) \) be the space of all sequences \( x = (x_j)_{j\in\mathbb{N}} \subset \mathbb{R} \) such that \( \|x\|_{\ell^2(\mathbb{N})} = \sum_{j\in\mathbb{N}} x_j^2 < \infty \). Show that the operator

\[
L : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}), \quad (x_1,x_2,x_3,\ldots) \mapsto (x_1,x_2/2,x_3/3,\ldots)
\]

is bounded, linear, and injective, but \( \text{Im} L \) is not closed.

Exercise A.6.12 Let \( \ell^2(\mathbb{N}) \) be the space of all sequences \( x = (x_j)_{j\in\mathbb{N}} \subset \mathbb{R} \) such that \( \|x\|_{\ell^2(\mathbb{N})} = \sum_{j\in\mathbb{N}} x_j^2 < \infty \). Determine which of the following operators \( L_j : \ell^2(\mathbb{N}) \to \ell^2(\mathbb{N}), j = 1, 2, \ldots, 4 \), satisfy an inf-sup condition:

\[
L_1 x = (x_1,0,x_2,0,x_3,0,\ldots),
\]
\[
L_2 x = (x_1,x_3,x_5,\ldots),
\]
\[
L_3 x = (x_1-x_2,x_3-x_4,\ldots),
\]
\[
L_4 x = (|x_1|,|x_2|,\ldots).
\]

Specify for each operator its adjoint.

Exercise A.6.13 Let \( X \) be a Hilbert space and let \( a : X \times X \to \mathbb{R} \) be a bounded and coercive bilinear form on \( X \). Show that \( a \) satisfies an inf-sup condition, and that \( a \) is nondegenerate.
Exercise A.6.14

(i) Show that \( p \in H^1_0(\Omega) \) solves the Poisson problem \( -\Delta p = g \) in \( \Omega \) with \( p|_{\partial \Omega} = 0 \) in the weak sense, if and only if the pair \((u, p) \in L^2(\Omega; \mathbb{R}^d) \times H^1_0(\Omega)\) with \( u = \nabla p \) satisfies

\[
\int_\Omega u \cdot v \, dx + \int_\Omega u \cdot \nabla p \, dx = 0,
\]

\[
- \int_\Omega u \cdot \nabla q \, dx = - \int_\Omega g q \, dx,
\]

for all \((v, q) \in L^2(\Omega; \mathbb{R}^d) \times H^1_0(\Omega)\).

(ii) Verify directly the boundedness, and the inf-sup and nondegeneracy conditions for the associated bilinear form

\[
\Gamma((u, p), (v, q)) = \int_\Omega u \cdot \nabla v \, dx + \int_\Omega u \cdot \nabla p \, dx + \int_\Omega u \cdot \nabla q \, dx.
\]

Conclude that the conditions of the generalized Lax–Milgram lemma are satisfied.

*Hint:* Consider the pair \((v, q) = (u - \nabla p, -2p)\).

Exercise A.6.15 Assume that \( V \) and \( Q \) are Hilbert spaces, and \( a : V \times V \to \mathbb{R} \) and \( b : V \times Q \to \mathbb{R} \) are bounded bilinear forms. Assume that the operator

\[
L : V \times Q \to V' \times Q', \quad (u, p) \mapsto (a(u, \cdot) + b(\cdot, p), b(u, \cdot))
\]

is an isomorphism, and there exists \( \beta > 0 \) such that

\[
\inf_{q \in Q' \setminus \{0\}} \sup_{v \in V \setminus \{0\}} \frac{b(v, q)}{\|v\| \|q\|_Q} \geq \beta.
\]

Show that there exists \( \alpha > 0 \) such that

\[
\inf_{v \in \ker B' \setminus \{0\}} \sup_{w \in \ker B' \setminus \{0\}} \frac{a(v, w)}{\|v\| \|w\|_V} \geq \alpha.
\]

Exercise A.6.16 Let \( a : V \times V \to \mathbb{R} \) be symmetric and positive semidefinite, and \( K \subset V \) a closed subspace. Assume that for every \( \ell \in K' \), there exists a unique \( u \in K \) such that

\[
a(u, v) = \ell(v)
\]

for all \( v \in K \) and such that \( \|u\|_V \leq c_K \|\ell\|_{K'} \).
(i) Prove the Cauchy–Schwarz inequality \( a(u, v)a(u, v) \leq a(u, u)a(v, v) \).

(ii) Show that \( a \) is coercive on \( K \).

**Exercise A.6.17** Let the bilinear forms \( a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R} \) and \( b : H^1(\Omega) \times \mathbb{R} \to \mathbb{R} \) be defined by

\[
a(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx, \quad b(v, \lambda) = \lambda \int_\Omega v \, dx.
\]

(i) Show that for every \( \ell \in H^1(\Omega)' \) there exists a uniquely defined pair \( (u, \lambda) \in H^1(\Omega) \times \mathbb{R} \) such that

\[
a(u, v) + b(v, \lambda) = \ell(v),
\]

\[
b(u, \mu) = 0,
\]

for all \( (v, \mu) \in H^1(\Omega) \times \mathbb{R} \).

(ii) Show that under appropriate conditions on functions \( f \in L^2(\Omega) \) and \( g \in L^2(\partial \Omega) \), and an appropriate definition of \( \ell \), the weak formulation defines a weak solution for the Poisson problem \(-\Delta u = f\) in \( \Omega \) with Neumann boundary condition \( \partial_n u = g \) on \( \Gamma_N = \partial \Omega \).

**Exercise A.6.18**

(i) Show that for all \( v \in H^1_0(\Omega) \), we have

\[
\frac{1}{2} \int_\Omega |\nabla v|^2 \, dx = \sup_{q \in L^2(\Omega; \mathbb{R}^d)} \int_\Omega q \cdot \nabla u \, dx - \frac{1}{2} \int_\Omega |q|^2 \, dx.
\]

(ii) Let \( f \in L^2(\Omega) \). Derive the optimality conditions for the saddle-point problem

\[
\inf_{v \in H^1_0(\Omega)} \sup_{q \in L^2(\Omega; \mathbb{R}^d)} L(v, q), \quad L(v, q) = \int_\Omega q \cdot \nabla v \, dx - \frac{1}{2} \int_\Omega |q|^2 \, dx - \int_\Omega fv \, dx,
\]

i.e., compute the derivatives of the mappings \( t \mapsto L(v+tw, q) \) and \( t \mapsto L(v, q+ts) \) at \( t = 0 \).

(iii) Determine a partial differential equation that is satisfied by the element \( u \) of a saddle point \( (u, p) \).

**Exercise A.6.19** Let \( A \in \mathbb{R}^{n_A \times n_A} \) and \( B \in \mathbb{R}^{n_B \times n_A} \) with \( n_A, n_B \in \mathbb{N} \) such that \( n_A \geq n_B \). Assume that \( \dim \text{Im } B = n_B \) and \( v^T A w > 0 \) for all \( v, w \in \ker A \). Let \( C \in \mathbb{R}^{n_B \times n_B} \) be positive semidefinite and \( t \geq 0 \). Show that the matrix

\[
\begin{bmatrix}
A & B^T \\
B & -tC
\end{bmatrix}
\]

is regular.
Exercise A.6.20

(i) Assume that \(k u k V C k p k Q > \frac{1}{2}\)
and \(k u k V C k q k Q C t q j c / D C 4.2\). Show that
\(k u k V C k q k Q C t q j c / D C 4.2\).
(ii) Show that for \(x, y, z \geq 0\) with \(x > 0\) and \(0 < x \leq y + z\),
we have \(x \leq y^2/x + z\).

Quiz A.6.2
Decide for each of the following statements whether it is true or false.
You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>If (L : X \rightarrow X') is self-adjoint, i.e., (L' = L), then the inf-sup condition implies nondegeneracy.</td>
<td>True</td>
</tr>
<tr>
<td>If the image of a bounded linear operator is finite-dimensional, then it is closed.</td>
<td>True</td>
</tr>
<tr>
<td>If (X = Y) and (\Gamma : X \times Y \rightarrow \mathbb{R}) is positive semidefinite and satisfies the conditions of the generalized Lax–Milgram lemma, then (\Gamma) is coercive.</td>
<td>True</td>
</tr>
<tr>
<td>If (L) satisfies an inf-sup condition, then (L') is surjective.</td>
<td>True</td>
</tr>
<tr>
<td>If (L : X \rightarrow Y') is bijective and satisfies an inf-sup condition, then it is an isomorphism.</td>
<td>True</td>
</tr>
</tbody>
</table>

A.6.3 Approximation of Saddle-Point Problems

Exercise A.6.21 Assume that the bilinear forms \(a : V \times V \rightarrow \mathbb{R}\) and \(b : V \times Q \rightarrow \mathbb{R}\), and the families of subspaces \((V_h)_{h>0}\) and \((Q_h)_{h>0}\) satisfy the Babuška–Brezzi conditions. Suppose that for \(Kh = \{v_h \in V_h : b(v_h, q_h) = 0\text{ for all }q_h \in Q_h\}\),
we have \(Kh \subseteq \ker B\), where \(B : V \rightarrow Q'\) is defined by \(Bv = b(v, \cdot)\) for all \(v \in V\).
Show that for the approximation error \(u - u_h\), we have
\[
\|u - u_h\|_V \leq (k_a/\alpha) \inf_{v_h \in Kh} \|u - u_h - v_h\|_V.
\]

Exercise A.6.22 For \(n \in \mathbb{N}\) and \(h = 1/n\), let \(\mathcal{T}_h\) be the triangulation of \(\Omega = (0, 1)^2\)
consisting of halved squares of length side \(h\), with diagonals parallel to the vector \((1, 1)\).

(i) Show that if \(v_h \in \mathcal{S}_0^1(\mathcal{T}_h)^2\) satisfies \(\text{div } v_h = 0\), then we have \(v_h = 0\).
(ii) Show that for every \(h = 1/n > 0\), there exists a constant \(c_h > 0\) such that
\[
\|\text{div } v_h\|_{L^2(\Omega)} \geq c_h \|v_h\|_{L^2(\Omega)}
\]
for all \(v_h \in \mathcal{S}_0^1(\mathcal{T}_h)^2\). Prove that \(c_h \leq ch\) with a constant \(c > 0\) that is independent of \(h\).
(iii) Let \(u(x, y) = [\sin(x) \sin(y), \cos(x) \cos(y)]^T\) for \((x, y) \in T\), where \(T\) is for \(h > 0\) defined by \(T = \text{conv}\{(0, 0), (h, 0), (0, h)\}\). Show that \(\text{div } u = 0\),
\[
\text{div } \mathcal{J}_h u \approx h/2 \text{ for small } h, \text{ and } |D^2 u| \geq 1/2. \text{ Conclude that } \| \text{div}(u - \mathcal{J}_h u) \|_{L^2(T)} \geq c h \| D^2 u \|_{L^2(T)}.
\]

**Exercise A.6.23** Let \( a : V \times V \rightarrow \mathbb{R} \) be a bounded, coercive, and symmetric bilinear form, and let \( B : V \rightarrow Q \) be a bounded linear operator for Hilbert spaces \( V \) and \( Q \). Let \( 0 < t \leq 1 \) and \( \ell \in V' \), and let \( u \in V \) be such that

\[
a(u, v) + t^2(Bu, Bv) = \ell(v)
\]

for all \( v \in V \). For a family of subspaces \((V_h)_{h>0}\), let \((u_h)_{h>0} \subset V\) be the sequence of corresponding Galerkin approximations. Specify and discuss the dependence of the constant in Céa’s lemma on the parameter \( t > 0 \).

**Exercise A.6.24** Let \( a : V \times V \rightarrow \mathbb{R} \) be a coercive, bounded and symmetric bilinear form, and let \( b : V \times Q \rightarrow \mathbb{R} \) be a bounded bilinear form, so that the conditions of Brezzi’s splitting theorem are satisfied. Let \( c : Q \times Q \rightarrow \mathbb{R} \) be a bounded and positive semidefinite bilinear form, and let \( 0 < t \ll 1 \). Let \((V_h)_{h>0}\) and \((Q_h)_{h>0}\) be families of subspaces, such that the Babuška–Brezzi conditions are satisfied. Use the stability results about perturbed saddle-point problems to derive an error estimate for its numerical approximation that is independent of \( 0 < t \leq 1 \).

**Exercise A.6.25** Let \( \mathcal{T}_h \) be the triangulation of \( \Omega = (0, 1)^2 \) with nodes \( z_0 = (0, 0), z_1 = (1, 0), z_2 = (1, 1), z_3 = (0, 1), \) and \( z_5 = (1, 1)/2 \). Construct a function \( q_h \in \mathcal{L}^0(\mathcal{T}_h) \) with \( \int_{\Omega} q_h \, dx = 0 \), such that \( \text{div} v_h \neq q_h \) for all \( v_h \in \mathcal{S}^1_0(\mathcal{T}_h)^2 \), and conclude that these spaces do not lead to an inf-sup condition for

\[
b(v_h, q_h) = \int_{\Omega} q_h \text{div} v_h \, dx.
\]

**Exercise A.6.26** Specify a Fortin interpolant for the bilinear form

\[
b(v, q) = \int_{\Omega} v \cdot \nabla q \, dx
\]

on \( V \times Q \) with \( V = L^2(\Omega; \mathbb{R}^d) \) and \( Q = H^1_0(\Omega) \), and for the subspaces \( V_h = \mathcal{L}^0(\mathcal{T}_h)^d \) and \( Q_h = \mathcal{S}^1_0(\mathcal{T}_h) \) for a regular family of triangulations \((\mathcal{T}_h)_{h>0}\) of \( \Omega \).

**Exercise A.6.27** Let \( A \in \mathbb{R}^{n_A \times n_A}, B \in \mathbb{R}^{n_B \times n_A}, f \in \mathbb{R}^{n_A}, \) and \( g \in \mathbb{R}^{n_B} \). Show that \( x \in \mathbb{R}^{n_A} \) satisfies \( Bx = g \) and \( y^\top Ax = y^\top f \) for all \( y \in \mathbb{R}^{n_A} \) with \( By = 0 \), if and only if \( Bx = g \) and there exists \( \lambda \in \mathbb{R}^{n_B} \) such that \( Ax + B^\top \lambda = b \).

**Exercise A.6.28** Let \( X = Y = H^1_0(\Omega), f \in L^2(\Omega), \) and \( X_h = Y_h = \mathcal{S}^1_0(\mathcal{T}_h) \) for a regular family of triangulations \((\mathcal{T}_h)_{h>0}\) of \( \Omega \). Let

\[
\Gamma(u, v) = \int_{\Omega} \nabla u \cdot \nabla v \, dx, \quad \ell(v) = \int_{\Omega} fv \, dx.
\]

Prove that the conditions of the generalized Céa lemma are satisfied.
Exercise A.6.29 Let \((\mathcal{T}_h)_{h>0}\) be a regular family of triangulations of \(\Omega\) and let \(g \in L^2(\Omega)\). Let \(p \in H^1_0(\Omega)\) and \((\hat{p}_h)_{h>0}\) be the exact solution and its Galerkin approximations of the problem

\[-\Delta p = g \text{ in } \Omega, \quad p|_{\partial\Omega} = 0.\]

For every \(h > 0\), let \((u_h, p_h) \in L^0(\mathcal{T}_h)^d \times \mathcal{S}_0^1(\mathcal{T}_h)\) be the solution of the saddle-point formulation

\[
\begin{align*}
\int_{\Omega} u_h \cdot v_h \, dx - \int_{\Omega} v_h \cdot \nabla p_h \, dx &= 0, \\
\int_{\Omega} u_h \cdot \nabla q_h \, dx &= \int_{\Omega} g q_h \, dx,
\end{align*}
\]

for all \((v_h, q_h) \in L^0(\mathcal{T}_h)^d \times \mathcal{S}_0^1(\mathcal{T}_h)\). Show that for every \(h > 0\) we have \(\hat{p}_h = p_h\).

Exercise A.6.30 Let \(\mathcal{T}_h\) be the partition of the interval \((0, 1)\) defined by the nodes \(x_j = jh = j/N, j = 0, 1, \ldots, N\), and let \(\mathcal{S}^{p,k}(\mathcal{T}_h) \subset C^k([0, 1])\) be the space of spline functions with piecewise polynomial degree \(p \geq 0\) and differentiability order \(k \geq 0\). Let \(\mathcal{S}^{r,k}_0(\mathcal{T}_h)\) be the subspace of functions that vanish for \(x \in \{0, 1\}\). Determine pairs \((r, k)\) and \((s, \ell)\) such that an inf-sup condition

\[
\sup_{v_h \in \mathcal{S}^{r,k}_0(\mathcal{T}_h) \setminus \{0\}} \frac{\int_0^1 v_h' q_h \, dx}{\|v_h'\|} \geq \beta_h \|q_h\|
\]

holds for all \(q_h \in \mathcal{S}^{s,\ell}(\mathcal{T}_h)\) with a positive constant \(\beta_h > 0\).

Quiz A.6.3 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Unique solvability of (\Gamma(x_h, y_h) = \ell(y_h)) requires that (\dim X_h = \dim Y_h)</td>
<td></td>
</tr>
<tr>
<td>The inf-sup constant (\gamma \geq 0) for a bilinear form (\Gamma) is bounded from below by the inverse of the continuity constant (k^{-1}_F)</td>
<td></td>
</tr>
<tr>
<td>Uniformity of a discrete inf-sup condition means that the constants (\gamma_h) are positive for every (h &gt; 0)</td>
<td></td>
</tr>
<tr>
<td>In the case of the Stokes equation with (V_h = \mathcal{S}^{1,0}_0(\mathcal{T}_h)^d) and (Q_h = L^0(\mathcal{T}_h)), we have (K_h = {v_h \in V_h : \text{div } v_h = 0})</td>
<td></td>
</tr>
<tr>
<td>The locking effect refers to a limited flexibility of a finite element space for a particular problem</td>
<td></td>
</tr>
</tbody>
</table>
A.6 Saddle-Point Problems

A.6.4 Projects

Project A.6.1 Discretize the mixed formulation of the one-dimensional Poisson problem $-p'' = f$ in $\Omega = (0, 1)$, $p(0) = p(1) = 0$, i.e., the problem of determining $(u, p) \in H^1(\Omega) \times L^2(\Omega)$ such that

$$
\int_0^1 u v' \, dx + \int_0^1 p v' \, dx = 0,
$$

$$
\int_0^1 q u' \, dx = - \int_\Omega f q \, dx,
$$

for all $(v, q) \in H^1(\Omega) \times L^2(\Omega)$. Use the finite element space $S^1(\mathcal{T}_h)$ for the approximation of $u$ and the spaces $S^1(\mathcal{T}_h)$ and $L^0(\mathcal{T}_h)$ for the approximation of $p$. Compare the results for the case $f = 1$ with exact solution $p(x) = x(1-x)/2$.

Project A.6.2 Solutions of the Neumann problem $-\Delta u = f$ in $\Omega$, $\partial_n u = g$ on $\Gamma_N = \partial \Omega$ only exist if the compatibility condition

$$
\int_\Omega f \, dx + \int_{\Gamma_N} g \, ds = 0
$$

is satisfied, and are unique only up to a constant. Determine different finite element approximations using the normalizations

$$
u_h(z_0) = 0, \quad \int_\Omega u_h \, dx = 0,$$

and via minimizing the functional

$$
\tilde{E}_h(v_h) = \frac{1}{2} \int_\Omega |\nabla v_h|^2 \, dx - \int_\Omega f v_h \, dx - \int_{\Gamma_N} g v_h \, ds + \left( \int_\Omega v_h \, dx \right)^2.
$$

Implement the three approaches, discuss relations between them, and comment on the linear systems of equations. Test your implementations for the case $\Omega = (0, 1)^2$, $f = 1$, and $g = -1/4$.

Project A.6.3 Consider the Neumann problem $-\Delta u = f$ in $\Omega$, $\partial_n u = g$ on $\Gamma_N = \partial \Omega$, assume that the compatibility condition

$$
\int_\Omega f \, dx + \int_{\Gamma_N} g \, ds = 0
$$


is satisfied, and let \( u \in H^1(\Omega) \) be the unique solution of the problem satisfying
\[
\int_{\Omega} u \, dx = 0.
\]
Characterize \( u \) as the solution of a saddle-point problem, define a discrete saddle-point problem, and compare its solution using the backslash operator and the Uzawa algorithm for the case \( \Omega = (0, 1)^3, f = 1, g = -1/6. \)

**Project A.6.4** Consider the problem

\[
\Delta^2 u = f \quad \text{in } \Omega, \quad u = \Delta u = 0 \quad \text{on } \partial \Omega.
\]
Introduce the variable \( \xi = -\Delta u \), formulate a saddle-point problem, and approximate it with a \( P1-P1 \) finite element method. Test your problem for the cases \( \Omega = (0, 1)^2, f = 1, \) and \( \Omega = B_1(0). \) Identify the exact solution in the second case using that \( \Delta u = \partial_\tau^2 u + r^1 \partial_\tau u \) for a rotationally symmetric function \( u. \)

**Project A.6.5** Determine a function \( f \in L^2(\Omega; \mathbb{R}^2) \) with \( \Omega = (-1, 1)^2 \) so that the exact solution \( u \in H_0^1(\Omega; \mathbb{R}^2) \) of the variational formulation

\[
\int_{\Omega} \nabla u : \nabla v \, dx + \varepsilon^{-2} \int_{\Omega} \text{div} u \text{div} v \, dx = \int_{\Omega} f \cdot v \, dx
\]
for all \( v \in H_0^1(\Omega; \mathbb{R}^2) \) is for every \( \varepsilon > 0 \) given by

\[
u(x_1, x_2) = \begin{bmatrix}
\sin(2\pi x_2) \sin^2(\pi x_1) \\
-\sin(2\pi x_1) \sin^2(\pi x_2)
\end{bmatrix}.
\]
Investigate the approximation of the problem in \( \mathcal{S}_0^1(\mathcal{T}_h)^2 \) for sequences of triangulations and the parameters \( \varepsilon = 10^{-j}, j = 1, 2. \)

**Project A.6.6** Implement the discretization of the Stokes problem

\[
-\Delta u + \nabla p = f, \\
\text{div } u = 0,
\]
with \( P1 \)-finite elements for approximating \( u \) and \( p. \) Test it for \( \Omega = (-1, 1)^2, \Gamma_D = \partial \Omega, \) and

\[
u(x_1, x_2) = \pi \begin{bmatrix}
\sin(2\pi x_2) \sin^2(\pi x_1) \\
-\sin(2\pi x_1) \sin^2(\pi x_2)
\end{bmatrix},
\]

\[
p(x_1, x_2) = \cos(\pi x_1) \sin(\pi x_2).
\]
A.7 Mixed and Nonstandard Methods

Project A.6.7 For a compact $C^2$-submanifold $M \subset \mathbb{R}^3$, the nearest-neighbor projection $\pi_M(z)$ of a point $z \in \mathbb{R}^3$ onto $M$ is defined as a point $x \in M$ with $|z - x| = \min_{y \in M} |z - y|$. One can show that there exists an open neighborhood of $M$ in which $\pi_M$ is well defined. We assume that $M = f^{-1}(\{0\})$ and characterize the nearest-neighbor projection of $z$ as a saddle-point for the functional

$$G(x, \lambda) = \frac{1}{2} |x - z|^2 + \lambda f(x),$$

i.e., a solution $(x, \lambda)$ of the equation $F(x, \lambda) = 0$ with $F : \mathbb{R}^4 \rightarrow \mathbb{R}^4$ defined by

$$F(x, \lambda) = \begin{bmatrix} x - z + \lambda \nabla f(x) = 0 \\ f(x) \end{bmatrix}.$$ 

Formulate the Newton iteration for the solution of this equation and discuss its well-posedness and convergence. Implement and test it for cases of a sphere and torus.

Project A.6.8 Implement a $P_0$-$P_1$ method for the primal mixed formulation of the Poisson problem and verify experimentally that it coincides with the standard $P_1$ finite element method.

### A.7 Mixed and Nonstandard Methods

#### A.7.1 Mixed Methods for the Poisson Problem

**Exercise A.7.1** Prove that the space $H_N(\text{div}; \Omega)$ is a Hilbert space.

**Exercise A.7.2** Let $u \in L^2(\Omega; \mathbb{R}^d)$ be such that $u|_{\Omega_i} \in C^1(\overline{\Omega_i}; \mathbb{R}^d)$, $i = 1, 2, \ldots, I$, for a partition $(\Omega_i)_{i=1,\ldots,I}$ of $\Omega$. Prove that we have $u \in H(\text{div}; \Omega)$ if and only if $u|_{\Omega_i} \cdot n_i = -u|_{\Omega_j} \cdot n_j$ on every interface $\Gamma_{ij} = \partial \Omega_i \cap \partial \Omega_j$ with the outer unit normals $n_i$ and $n_j$ to $\Omega_i$ and $\Omega_j$.

**Exercise A.7.3** Let $\Omega = \{(x_1, x_2) \in \mathbb{R}^2 : |x| < 1, x_1, x_2 > 0\}$, $S = [0, 1] \times \{0\}$ and

$$u_\varepsilon : \Omega \rightarrow \mathbb{R}^2, \quad x \mapsto \frac{[-x_2, x_1]^T}{\varepsilon + |x|^2}.$$

(i) Show that $u_\varepsilon \in H(\text{div}; \Omega)$ and that there exists $\alpha_\varepsilon \in \mathbb{R}$ such that $\tilde{u}_\varepsilon = \alpha_\varepsilon u_\varepsilon \in H(\text{div}; \Omega)$ and $\|\tilde{u}_\varepsilon\|_{L^2(\Omega)} = 1$. 


(ii) Show that \( \| \tilde{u}_\varepsilon \cdot n \|_{L^1(S)} \) is unbounded as \( \varepsilon \to 0 \), and conclude that the trace operator for functions in \( H(\text{div}; \Omega) \) is not well-defined as an operator into \( L^1(\partial \Omega) \) in general.

(iii) Why is the expression \( \int_{\partial \Omega} u_\varepsilon \cdot n \, ds \) well-defined?

**Exercise A.7.4** Let \( p_D \in H^1(\Omega) \), \( g \in L^2(\Omega) \), and \( \sigma \in L^2(\Gamma_N) \). Let \( p \in H^1(\Omega) \) be the weak solution of the Poisson problem, i.e., we have \( p = p_D \) on \( \Gamma_D \) and

\[
\int_{\Omega} \nabla p \cdot \nabla q \, dx = \int_{\Omega} g q \, dx + \int_{\Gamma_N} \sigma q \, ds
\]

for all \( q \in H^1_0(\Omega) \). Let \( (u, p') \in H(\text{div}; \Omega) \times L^2(\Omega) \) be the solution of the dual mixed formulation of the Poisson problem, i.e., we have \( u \cdot n = \sigma \) on \( \Gamma_N \) and

\[
\int_{\Omega} u \cdot v \, dx + \int (p' \text{div} v) \, dx = \langle v \cdot n, p_D \rangle_{\partial \Omega},
\]

\[
\int_{\Omega} q \text{div} u \, dx = -\int_{\Omega} g q \, dx,
\]

for all \( (v, q) \in H_N(\text{div}; \Omega) \times L^2(\Omega) \). Prove that \( u = \nabla p \) and \( p = p' \) and conclude that the dual mixed formulation is well-posed.

**Exercise A.7.5** Let \( T \subset \mathbb{R}^d, d = 1, 2, 3 \), be a nondegenerate simplex. Let \( S \subset \partial T \) be a side of \( T \) with outer unit normal \( n_S \), and let \( z \in T \) be the vertex of \( T \) opposite \( S \). Prove that

\[
d!! |T| = |S|(z - x_S) \cdot n_S
\]

for an arbitrary point \( x_S \in S \).

**Exercise A.7.6** For \( v \in H(\text{div}; T) \) and an affine diffeomorphism \( \Phi_T : \hat{T} \to T \), define

\[
\hat{v} = (\det D\Phi_T)(D\Phi_T)^{-1}(v \circ \Phi_T).
\]

Compute the divergence of \( \hat{v} \), and show that for every side \( \hat{S} \subset \partial \hat{T} \) and \( S = \Phi_T(\hat{S}) \) we have

\[
\int_{S} v \cdot n \, ds = \int_{\hat{S}} \hat{v} \cdot \hat{n} \, d\hat{s}.
\]
Exercise A.7.7 Let $v \in H^1(\omega)$ and define $\overline{v} = |\omega|^{-1} \int_\omega v \, dx$.

(i) Show that $\|\overline{v}\|_{L^2(\omega)} \leq \|v\|_{L^2(\omega)}$.

(ii) Prove that

\[ \|v - \overline{v}\|_{L^2(\omega)} \leq c \text{diam}(\omega) \|\nabla v\|_{L^2(\omega)}, \]

with a constant $c > 0$ that is independent of $v$ and diam$(\omega)$.

Exercise A.7.8 Let $T \subset \mathbb{R}^d$ be a simplex, and let $\widehat{T} \subset \mathbb{R}^d$ be a reference simplex. Let $S$ and $\widehat{S}$ be sides of $T$ and $\widehat{T}$, respectively.

(i) Prove that there exists a constant $c > 0$, such that for all $\widehat{v} \in H^1(\widehat{T})$ we have

\[ \|\widehat{v}\|_{L^2(\widehat{S})} \leq c \|\widehat{v}\|_{H^1(\widehat{T})}. \]

(ii) Prove that there exists a constant $c > 0$, such that for all $v \in H^1(T)$ we have

\[ \|v\|_{L^2(S)} \leq c (h_T^{-1/2} \|v\|_{H^1(T)} + h_T^{1/2} \|\nabla v\|_{L^2(T)}). \]

Exercise A.7.9 Assume that $\Omega \subset \mathbb{R}^d$, $I_D \subset \partial \Omega$, and $I_N = \partial \Omega \setminus I_D$ are such that the Poisson problem

\[ -\Delta \phi = q \text{ in } \Omega, \quad \partial_n \phi|_{I_N} = 0, \quad \phi|_{I_D} = 0 \]

is $H^2$-regular, i.e., there exists $c > 0$ such that $\|D^2 \phi\|_{L^2(\Omega)} \leq c \|q\|_{L^2(\Omega)}$ for every $q \in L^2(\Omega)$. Use the Fortin interpolant $\mathcal{I}_{\mathcal{T}} : H^1(\Omega; \mathbb{R}^d) \to \mathcal{T}_N(\mathcal{T}_h)$ to prove the discrete inf-sup condition for the bilinear form $b$ in the dual mixed formulation of the Poisson problem.

Exercise A.7.10 Let $u \in H(\text{div}; \Omega)$. Show that there exists a sequence $(u_\varepsilon)_{\varepsilon > 0} \subset C^\infty(\overline{\Omega}; \mathbb{R}^d)$ such that

\[ \|u - u_\varepsilon\|_{H(\text{div};\Omega)} \leq c_\varepsilon, \quad \|\nabla u_\varepsilon\|_{L^2(\Omega)} \leq c \varepsilon^{-1} \]

with $c_\varepsilon \to 0$ as $\varepsilon \to 0$. 
**Quiz A.7.1** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True or False</th>
</tr>
</thead>
<tbody>
<tr>
<td>The mapping ((u, q) \mapsto \langle u \cdot n, q \rangle_{\partial \Omega}) defines a bounded bilinear form on (H(\text{div}; \Omega) \times H^1(\Omega))</td>
<td>True</td>
</tr>
<tr>
<td>A piecewise polynomial vector field belongs to (H^1(\Omega; \mathbb{R}^d)) if its tangential component is continuous across interfaces</td>
<td>True</td>
</tr>
<tr>
<td>The Raviart–Thomas finite element space consists of all piecewise affine vector fields with continuous normal components across sides</td>
<td>True</td>
</tr>
<tr>
<td>The divergence operator defines a bijection between the spaces (\mathcal{AT}^0(\mathcal{T}_h)) and (L^0(\mathcal{T}_h))</td>
<td>True</td>
</tr>
<tr>
<td>There exists a bounded linear Fortin interpolant (I_{\mathcal{T}} : H_N(\text{div}; \Omega) \rightarrow \mathcal{AT}^0_N(\mathcal{T}_h)) for the dual mixed Poisson problem</td>
<td>True</td>
</tr>
</tbody>
</table>

### A.7.2 Approximation of the Stokes System

**Exercise A.7.11** Show that the space \(L^2_0(\Omega) = \{ q \in L^2(\Omega) : \int_\Omega q \, dx = 0 \}\) is a Hilbert space which is isomorphic to the quotient space \(L^2(\Omega)/\mathbb{R}\), resulting from identifying functions that coincide up to an additive constant, and which is equipped with the quotient space norm

\[
\|q\|_{L^2(\Omega)/\mathbb{R}} = \inf_{c \in \mathbb{R}} \|q - c\|_{L^2(\Omega)}.
\]

**Exercise A.7.12** Specify the linear system of equations resulting from the discretization of the Stokes system with the MINI-element, and show that the degrees of freedom related to bubble functions can be eliminated by inverting a diagonal matrix.

**Exercise A.7.13** Let \(\mathcal{T}_h\) be a triangulation of \(\Omega \subset \mathbb{R}^d\) with sides \(\mathcal{I}_h\). For \(q_h \in \mathcal{P}^{1,c}(\mathcal{T}_h)\) and an inner side \(S \in \mathcal{I}_h\), let \([q_h]|_S\) be the jump of \(q_h\) across \(S\) defined by

\[
[q_h(x)] = \lim_{\varepsilon \to 0} \left( q_h(x + \varepsilon n_S) - q_h(x - \varepsilon n_S) \right)
\]

for every \(x\) in the interior of \(S\).

(i) Show that for every \(q_h \in \mathcal{P}^{1,c}(\mathcal{T}_h)\) and every \(S \in \mathcal{I}_h\) we have

\[
\int_S [q_h] \, ds = 0.
\]
(ii) Let \( u_h \in \mathcal{R}^0(\mathcal{T}_h) \) and \( q_h \in \mathcal{S}^1(\mathcal{T}_h) \). Show that
\[
\int_{\Omega} q_h \text{div} u_h \, dx = -\int_{\Omega} \nabla q_h \cdot u_h \, dx + \int_{\partial\Omega} q_h u_h \cdot n \, ds.
\]

**Exercise A.7.14** Let \( \mathcal{T}_h \) be a triangulation of \( \Omega \subset \mathbb{R}^d \).

(i) For \( u \in H^1_0(\Omega) \) let \( G_h u \in \mathcal{S}^1(\mathcal{T}_h) \) be defined by
\[
\int_{\Omega} \nabla G_h u \cdot \nabla v_h \, dx = \int_{\Omega} \nabla u \cdot \nabla v_h \, dx
\]
for all \( v_h \in \mathcal{S}^1(\mathcal{T}_h) \). Show that \( G_h u \) is well defined with \( \|\nabla G_h u\|_{L^2(\Omega)} \leq \|\nabla u\|_{L^2(\Omega)} \). Prove that if the Poisson problem is \( H^2 \)-regular in \( \Omega \), then we have
\[
\|u - G_h u\|_{L^2(\Omega)} \leq c_1 h \|\nabla u\|_{L^2(\Omega)}.
\]

(ii) For \( T \in \mathcal{T}_h \) with vertices \( z_1, z_2, \ldots, z_{d+1} \in \mathcal{N}_h \) let \( b_T = \varphi_{z_1} \varphi_{z_2} \cdots \varphi_{z_{d+1}} \). Given \( v \in L^2(\Omega) \) and \( T \in \mathcal{T}_h \), let \( \lambda_T \in \mathbb{R} \) be such that
\[
\int_T (\lambda_T b_T - v) \, dx = 0
\]
and define \( R_h v = \sum_{T \in \mathcal{T}_h} \lambda_T b_T \). Show that \( \|\nabla R_h v\|_{L^2(\Omega)} \leq c_2 h^{-1} \|v\|_{L^2(\Omega)} \) for all \( v \in L^2(\Omega) \), where \( h_{\min} = \min_{T \in \mathcal{T}_h} \text{diam}(T) \).

*Hint:* Show that \( \|\nabla b_{T_{\text{ref}}}\|_{L^2(T_{\text{ref}})} \leq c_2' \|b_{T_{\text{ref}}}\|_{L^2(T_{\text{ref}})} \) and use a transformation argument.

For \( w \in H^1_0(\Omega; \mathbb{R}^d) \) let \( I_F w = G_h w + R_h (w - G_h w) \), where \( G_h \) and \( R_h \) are applied component-wise. Show that there exists \( c > 0 \)
\[
\|\nabla I_F w\|_{L^2(\Omega)} \leq c (h_{\max}/h_{\min}) \|\nabla w\|_{L^2(\Omega)}.
\]

(iii) Show that for all \( w \in H^1_0(\Omega; \mathbb{R}^d) \) and all \( q_h \in \mathcal{S}^1(\mathcal{T}_h) \), we have
\[
\int_{\Omega} q_h \text{div}(w - I_F w) \, dx = 0.
\]

(iv) Formulate sufficient conditions for the uniform validity of the inf-sup condition for the spaces \( V_h = \mathcal{S}^1(\mathcal{T}_h)^d \oplus \mathcal{R}(\mathcal{T}_h)^d \) and \( Q_h = \mathcal{S}^1(\mathcal{T}_h) \cap L^2_0(\Omega) \) for the approximation of the Stokes system.
Exercise A.7.15

(i) Show that $v_h \mapsto \|\nabla_T v_h\|_{L^2(\Omega)}$ defines a norm on $\mathcal{S}_D^{1,cr}(\mathcal{T}_h)$.
(ii) Let $u_h \in \mathcal{S}_D^{1,cr}(\mathcal{T}_h)$. Show that in general $u_h|_{\Gamma_D} \neq 0$.
(iii) Show that there exists a uniquely defined function $u_h \in \mathcal{S}_D^{1,cr}(\mathcal{T}_h)$ with $u_h(x_S) = u_D(x_S)$ for all midpoints $x_S$ of sides $S \in \mathcal{T}_h \cap \Gamma_D$ and

$$
\int_{\Omega} \nabla_T u_h \cdot \nabla_T v_h \, dx = \int_{\Omega} f v_h \, dx + \int_{\Gamma_N} g v_h \, ds
$$

for all $v_h \in \mathcal{S}_D^{1,cr}(\mathcal{T}_h)$.
(iv) Show that for all $v_h \in \mathcal{S}_D^{1}(\mathcal{T}_h)$, we have

$$
\int_{\Omega} \nabla_T(u - u_h) \cdot \nabla(u - u_h - v_h) \, dx = 0,
$$

and conclude a best approximation result.
(v) Show that $\mathcal{S}_D^{1}(\mathcal{T}_h) \subset \mathcal{S}_D^{1,cr}(\mathcal{T}_h)$.

Exercise A.7.16 Let $\Gamma : X \times X \to \mathbb{R}$ be a symmetric bounded bilinear form that satisfies an inf-sup condition. For a subset $X_h \subset X$ and a bounded symmetric positive semidefinite bilinear form $c_h : X_h \times X_h \to \mathbb{R}$ let $\hat{\Gamma}_h = \Gamma + c_h$. Assume that $\hat{\Gamma}_h : X_h \times X_h \to \mathbb{R}$ satisfies an inf-sup condition uniformly in $h > 0$. For a given functional $\ell \in X'$, let $x \in X$ and $x_h \in X_h$ be the solutions of

$$
\Gamma(x, y) = \ell(y), \quad \hat{\Gamma}_h(x_h, y_h) = \ell(y_h)
$$

for all $y \in X$ and $y_h \in X_h$, respectively. Prove that

$$
c^{-1}\|x - x_h\| \leq \inf_{v_h \in X_h} \|x - v_h\|_X + c_h(v_h, v_h)^{1/2}.
$$

Exercise A.7.17 For $v \in L^1(\Omega)$, the distributional gradient is the operator

$$
\nabla v : C_0^\infty(\Omega; \mathbb{R}^d) \to \mathbb{R}, \quad \phi \mapsto -\int_\Omega v \, \text{div} \phi \, dx.
$$

Show that if $v \in H^1(\Omega)$, then the distributional gradient can be identified with the elementwise weak gradient $\nabla_T v$, but not if $v \in H^1(\mathcal{T}_h)$.

Exercise A.7.18

(i) Show that the Crouzeix–Raviart element is a finite element $(T, \mathcal{P}_1(T), \mathcal{K})$ for an appropriate choice of the functionals $\mathcal{K}$, and determine the corresponding nodal basis functions.
(ii) Show that the space $\mathcal{S}^{1,cr}(\mathcal{T}_h)$ is an affine family.

(iii) Prove that the Crouzeix–Raviart element is not a $C^0$-element.

**Exercise A.7.19** Let $u_h \in \mathcal{S}^{1,cr}(\mathcal{T}_h)$ be the Crouzeix–Raviart approximation of the Poisson problem $-\Delta u = f$ in $\Omega$, $u|_{\partial\Omega} = 0$, defined by

$$\int_{\Omega} \nabla \mathcal{S} u_h \cdot \nabla \mathcal{S} v_h \, dx = \int_{\Omega} f v_h \, dx$$

for all $v_h \in \mathcal{S}^{1,cr}(\mathcal{T}_h)$. Let $a_h(u_h, v_h)$ and $b_h(v_h)$ be defined by the left- and right-hand sides of the identity. Assume that the Poisson problem is $H^2$-regular and prove that

$$\sup_{v_h \in \mathcal{S}^{1,cr}(\mathcal{T}_h) \setminus \{0\}} \frac{|a_h(u, v_h) - b_h(v_h)|}{\|\nabla \mathcal{S} v_h\|} \leq c h \|D^2 u\|_{L^2(\Omega)}.$$

Deduce an error estimate with the second Strang lemma.

**Hint:** Use that $\int_S [v_h] \, ds = 0$ for every interior side $S \in \mathcal{T}_h$ and $v_h|_{T_{\pm}}(x) = v_h(x_S) + \nabla v_h|_{T_{\pm}} \cdot (x - x_S)$ on the neighboring elements $T_{\pm}$ to $S$ and with $x_S \in S$.

**Exercise A.7.20** Assume that $\mathcal{T}_h$ is a quasiuniform triangulation of $\Omega \subset \mathbb{R}^d$, and let $V_h = \mathcal{S}^k(\mathcal{T}_h)^d$ and $Q_h = \mathcal{S}^{k-1}(\mathcal{T}_h) \cap L^2(\Omega)$, for $k \geq 2$.

(i) Show that

$$\sup_{v_h \in V_h \setminus \{0\}} \frac{\int_{\Omega} q_h \, \text{div} \, v_h \, dx}{\|\nabla v_h\|_{L^2(\Omega)}} \geq c_1 h \int_{\Omega} \nabla q_h \cdot v_h \, dx \frac{\|v_h\|_{L^2(\Omega)}}{\|v_h\|_{L^2(\Omega)}}.$$

(ii) Use the Clément quasi-interpolant to prove that

$$\sup_{v_h \in V_h \setminus \{0\}} \frac{\int_{\Omega} q_h \, \text{div} \, v_h \, dx}{\|\nabla v_h\|_{L^2(\Omega)}} \geq \beta' \|q_h\|_{L^2(\Omega)} - c_2 h \|\nabla q_h\|_{L^2(\Omega)}.$$

(iii) Assume that for every $q_h \in Q_h$, there exists $v_h \in V_h$ such that

$$\int_{\Omega} \nabla q_h \cdot v_h \, dx \geq c_3 \|\nabla q_h\|_{L^2(\Omega)}^2, \quad \|v_h\|_{L^2(\Omega)} \leq c_4 \|\nabla q_h\|_{L^2(\Omega)}.$$

and deduce the inf-sup condition for the spaces $V_h$ and $Q_h$.

(iv) Try to construct $v_h \in V_h$ with the assumed properties.
Quiz A.7.2 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Truth</th>
</tr>
</thead>
<tbody>
<tr>
<td>The bilinear form $b$ satisfies an inf-sup condition if and only if $-b$ satisfies an inf-sup condition</td>
<td></td>
</tr>
<tr>
<td>The MINI-element satisfies an inf-sup condition if $\mathcal{D} \subset \mathbb{R}^2$ is convex and $\Gamma_D = \partial \mathcal{D}$</td>
<td></td>
</tr>
<tr>
<td>The $P_1$-$P_0$ method defines a nonconforming stable Stokes element</td>
<td></td>
</tr>
<tr>
<td>If a Crouzeix–Raviart function is continuous at two distinct points on a side of a triangle, then it is continuous across the side</td>
<td></td>
</tr>
<tr>
<td>We have $\dim \mathcal{X}^{1,cr}(\mathcal{T}_h) =</td>
<td>\mathcal{T}_h</td>
</tr>
</tbody>
</table>

A.7.3 Convection-Dominated Problems

Exercise A.7.21 Let $\Omega = (0, L_1) \times (0, L_2) \times (-L_3, L_3)$ and let $b = 10 \text{m/s} \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T$ be the velocity field of the wind in $\Omega$. Smoke is released from a chimney at $(0, \ell_{ch}, 0)$ with $\ell_{ch}$ being the height of the chimney, cf. Fig. A.7. Let $c : [0, T] \times \Omega \to \mathbb{R}$ be the smoke concentration with units $[c] = g/m^3$.

(i) Explain the following principle:

$$\int_\omega c(t + \delta t, x + b\delta t) \, dx = \int_\omega c(t, x) \, dx - \int_t^{t+\Delta t} \int_{\partial \omega} q(s) \cdot n(s) \, ds$$

for a control volume $\omega \subset \Omega$, $t > 0$ and $\delta t$ sufficiently small. Make use of Fick’s law, i.e., $q = -\nu \nabla c$, where $\nu = 1.5 \cdot 10^{-5} \text{m}^2/\text{s}$ is the diffusion coefficient for carbon dioxide in air.

(ii) Deduce a partial differential equation for $c$. Non-dimensionalize the equation and explain why the process is convection dominated.

(iii) Derive a two-dimensional simplification by introducing

$$\bar{c}(x_1, x_2) = \int_{-L_3}^{L_3} c(x_1, x_2, x_3) \, dx_3,$$

and formulate appropriate boundary conditions on $\partial \Omega$.

Fig. A.7 Smoke released from a chimney and transported by a wind field
(iv) Justify the assumption \( c(t + \delta t, x) = c(t, x) \) for \( x \in \Omega \) and \( t \in [0, T] \), and deduce a steady state convection-diffusion equation.

**Exercise A.7.22** Let \( \Omega = (0, 1)^2 \) and define \( b(x) = (\sin(\phi), -\cos(\phi)) \) for \( x = r(\cos(\phi), \sin(\phi)) \in \Omega \). Show that \( \text{div} \, b = 0 \) and construct a solution of the equation \( b \cdot \nabla u = 0 \) in \( \Omega \) subject to a boundary condition defined by \( u_D(x_1, x_2) = x_2 \) imposed on a suitable subset of \( \partial \Omega \).

**Exercise A.7.23** Let \( b \in H_N(\text{div}; \Omega) \cap L^\infty(\Omega; \mathbb{R}^d) \) with \( \text{div} \, b = 0 \). Show that the bilinear form

\[
c(u, v) = \int_\Omega b \cdot (\nabla u) v \, dx
\]

is bounded and skew-symmetric on \( H^1_D(\Omega) \).

**Exercise A.7.24** For \( b \in H_N(\text{div}; \Omega) \cap L^\infty(\Omega; \mathbb{R}^d) \) with \( \text{div} \, b = 0 \) and \( u_D \in C(\partial \Omega) \), we consider the equation

\[
-\varepsilon \Delta u + b \cdot \nabla u = 0 \text{ in } \Omega, \quad u|_{\partial \Omega} = u_D.
\]

(i) Prove by considering \( \tilde{u} = \min\{u, c\} \) for an appropriate choice of \( c \) that \( u(x) \leq \max_{y \in \partial \Omega} u_D(y) \) for \( x \in \Omega \).

(ii) Prove \( \|D^2 u\|_{L^2(\Omega)} \leq c\varepsilon^{-2} \) provided that the Poisson problem is \( H^2 \)-regular.

**Exercise A.7.25** Compare the error estimates for the standard Galerkin approximation and the streamline-diffusion method.

**Exercise A.7.26** For grid points \( x_i = i/M, \, i = 0, 1, \ldots, M \), and \( a, b \in \mathbb{R} \), consider the scheme

\[
-\varepsilon \frac{U_{i-1} - 2U_i + U_{i+1}}{h^2} + \frac{U_{i+1} - U_{i-1}}{2h} = 0, \quad U_0 = a, \quad U_M = b,
\]

for \( i = 1, 2, \ldots, M - 1 \) with \( h = 1/M \). Rewrite the scheme in the form \( \hat{U}_{i+1} = A \hat{U}_i \) with a matrix \( A \in \mathbb{R}^{2 \times 2} \) and \( \hat{U}_{i+1} = [U_{i+1}, U_i]^\top \), and construct the solution in terms of the eigenvalues of \( A \).

**Exercise A.7.27**

(i) Devise an upwinding finite difference discretization of the equation

\[
-\varepsilon u'' + bu' = 0, \quad u(0) = \alpha, \quad u(1) = \beta,
\]

and prove a discrete maximum principle.
(ii) Show that the upwinding discretization of the convection term $bu^\prime$ can be interpreted as a symmetric discretization of the modified term $bu^\prime + \delta u''$ with an appropriate parameter $\delta$.

**Exercise A.7.28** Assume that the nodes $z, y$ are the endpoints of an inner edge $S = T_1 \cap T_2$ in a Delaunay triangulation. Show that for the associated nodal basis functions $\varphi_z, \varphi_y \in \mathcal{S}^1(\mathcal{T}_h)$, we have

$$\int_{\Omega} \nabla \varphi_z \cdot \nabla \varphi_y \, dx = -\frac{|m_{T_1} - m_{T_2}|}{|z - y|} = -\frac{1}{2} (\cot \alpha_1 + \cot \alpha_2),$$

where $m_1, m_2$ are the circumcenters of $T_1, T_2$, and $\alpha_1, \alpha_2$ are the inner angles of $T_1, T_2$ that are opposite $S$.

**Exercise A.7.29**

(i) Construct the Voronoi diagram associated with the given points $(x_j)_{j=1,\ldots,8} \subset \overline{\Omega} = [0, 5] \times [0, 2]$ shown in Fig. A.8.

(ii) Discuss the regularity of the diagram.

(iii) Construct the Delaunay triangulation of the Voronoi diagram and verify its weak acuteness.

**Exercise A.7.30** Let $b \in C^1(\overline{\Omega}; \mathbb{R}^d)$ with $\text{div} \, b = 0$ and $\varepsilon, \alpha > 0$, and consider the boundary value problem

$$-\varepsilon \Delta u + b \cdot \nabla u + \alpha u = f \text{ in } \Omega, \quad u|_{\partial \Omega} = 0.$$

Show that for the $P_1$-finite element approximation of the problem, we have the error estimate

$$\varepsilon^{1/2} \|\nabla e\|_{L^2(\Omega)} + \alpha^{1/2} \|e\|_{L^2(\Omega)} \leq c(\varepsilon^{1/2} + \varepsilon^{-1/2} h \|b\|_{L^\infty(\Omega)} + \alpha^{1/2} h) \|D^2 u\|_{L^2(\Omega)}$$

for $e = u - u_h$. Discuss extreme parameters $\alpha, \varepsilon$ for which the Galerkin approximation provides useful approximations.

**Fig. A.8** Points $(x_j)_{j=1,\ldots,5}$ define a Voronoi diagram
Quiz A.7.3  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A convection-diffusion problem can be formulated as an equivalent minimization problem</td>
<td>True</td>
<td>Justify decision</td>
</tr>
<tr>
<td>Boundary layers occur when a Dirichlet boundary is imposed on an outflow boundary</td>
<td>True</td>
<td>Justify decision</td>
</tr>
<tr>
<td>The solution of a convection-dominated partial differential equation is uniformly bounded as to $\varepsilon \to 0$ in $H^1(\Omega)$ but not in $H^2(\Omega)$</td>
<td>False</td>
<td>Justify decision</td>
</tr>
<tr>
<td>The streamline-diffusion method finite element methods by modifying the diffusion term</td>
<td>True</td>
<td>Justify decision</td>
</tr>
<tr>
<td>The finite volume method allows for a generalization of upwinding techniques to higher-dimensional problems</td>
<td>True</td>
<td>Justify decision</td>
</tr>
</tbody>
</table>

A.7.4  **Discontinuous Galerkin Methods**

Exercise A.7.31  Let $\mathcal{T}_h$ be the triangulation of $\Omega = (-1, 1)^2$ consisting of four halved squares with diagonals parallel to the vector $(1, 1)$. Compute the jumps and averages of the function

$$v(x, y) = \begin{cases} 
   x^2y, & x \geq 0, y > 0, \\
   x(1-y), & x < 0, y \geq 0, \\
   x^3, & x \leq 0, y < 0, \\
   1-y, & x > 0, y \leq 0.
\end{cases}$$

Exercise A.7.32

(i) Prove that if $v \in H^1(\Omega)$, then for every interior side $S \in \mathcal{T}_h$ we have that

$$[v]_S = 0, \quad \{v\}_S = v|_S.$$

(ii) Show that if $v \in H^2(\mathcal{T}_h)$ such that $\nabla v \in H(\text{div}; \Omega)$, then

$$[\nabla v \cdot n_S]_S = 0,$$

where $n_S$ is extended constantly to a neighborhood of $S$.

Exercise A.7.33  Let $u_D = \tilde{u}_D$ for $\tilde{u}_D \in H^1(\Omega) \setminus H^2(\Omega)$. Devise a discontinuous Galerkin method for approximating the Poisson problem $-\Delta u = f$ in $\Omega$ with $u|_{\partial\Omega} = u_D$. 
Exercise A.7.34  Let $\beta_S > 0$ for every $S \in \mathcal{S}_h$. Prove that
\[
\|v\|_{dG}^2 = \|
abla \mathcal{T} v\|_{L^2(\Omega)}^2 + \sum_{S \in \mathcal{S}_h} \frac{\beta_S}{h_S^2} \int_S \|v\|^2 \, ds
\]
is a norm on $H^1(\mathcal{T}_h)$.

Exercise A.7.35

(i) Show that the bilinear form $a_{dG} : \mathcal{H}^k(\mathcal{T}_h) \times \mathcal{H}^k(\mathcal{T}_h) \to \mathbb{R}$ defined by
\[
a_{dG}(u_h, v_h) = \int_\Omega \nabla \mathcal{T} u_h \cdot \nabla \mathcal{T} v_h \, dx + \sum_{S \in \mathcal{S}_h} \int_S \{\nabla u_h \cdot n_S\}[v_h] \, ds + \sigma \sum_{S \in \mathcal{S}_h} \int_S \{\nabla v_h \cdot n_S\}[u_h] \, ds + \sum_{S \in \mathcal{S}_h} \frac{\beta_S}{h_S^2} \int_S [u_h][v_h] \, ds
\]
is bounded with respect to the norm $\| \cdot \|_{dG}$.

(ii) Show that $a_{dG}$ is symmetric if and only if $\sigma = 1$.

Exercise A.7.36  Show that $u \in H^1_0(\Omega) \cap H^2(\mathcal{T}_h)$ is a weak solution of the Poisson problem $-\Delta u = f$, if and only if
\[
a_{dG}(u, v) = \ell(v)
\]
for all $v \in H^2(\mathcal{T}_h)$ and $\nabla u \in H(\text{div}; \Omega)$.

Exercise A.7.37

(i) Show that for $v \in H^1(\mathcal{T}_h)$ and a side $S \subset \partial T$, we have
\[
\|v\|_{L^2(S)} \leq c_{\text{Tr}}(h^{1/2}_S \|
abla v\|_{L^2(T)} + h^{-1/2}_S \|v\|_{L^2(T)}).
\]

(ii) Prove that for $v_h \in \mathcal{H}^k(\mathcal{T}_h)$, we have
\[
\|v_h\|_{L^2(S)} \leq c_{\text{Tr}, k} h^{-1/2}_S \|v_h\|_{L^2(T)},
\]
where $c_{\text{Tr}, k} \to \infty$ as $k \to \infty$.

Exercise A.7.38  Assume that the Poisson problem with homogeneous Dirichlet conditions on $\partial \Omega$ is $H^2$-regular. Use the representation
\[
\|v\|_{L^2(\Omega)} = \sup_{q \in L^2(\Omega) \setminus \{0\}} \frac{\int_\Omega v q \, dx}{\|q\|_{L^2(\Omega)}}
\]
to prove that there exists $c_{P,dG} > 0$ such that

$$\|v_h\|_{L^2(\Omega)} \leq c_{P,dG} \|v_h\|_{dG}$$

for all $v_h \in S^{1,dG}(\mathcal{T}_h)$.

**Exercise A.7.39** Let $A_{dG}$ be the matrix representing the symmetric bilinear form $a_{dG}$ in the basis $(\varphi_{T,z} : T \in \mathcal{T}_h, z \in \mathcal{N}_h \cap T)$ with $\varphi_{T,z}$ defined with the $P_1$-hat functions by

$$\varphi_{T,z}(x) = \begin{cases} \varphi_z(x) & \text{if } x \in T, \\ 0 & \text{otherwise.} \end{cases}$$

Show that $\text{cond}_2(A_{dG}) = \mathcal{O}(h^{-1-\gamma})$.

**Exercise A.7.40** Let $\mathcal{T}_h$ be a fixed triangulation and let $u_{h}^\beta \in S^{1,dG}(\mathcal{T}_h)$ be the solution corresponding to a discontinuous Galerkin method for the Poisson problem with $\beta_s = \beta$ for all $S \in \mathcal{T}_h$. Show that as $\beta \to \infty$, the sequence $(u_{h}^\beta)_{\beta \geq 0}$ converges to a function $u_h \in S^1(\mathcal{T}_h)$ that is the continuous Galerkin approximation.

**Quiz A.7.4** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

If $v \in H^1(\mathcal{T}_h)$ with $[v]_S = 0$ for all $S \in \mathcal{T}_h$, then $v \in H^1_0(\Omega)$

- The symmetric interior penalty discontinuous Galerkin method is unconditionally well-posed

- The nonsymmetric interior penalty discontinuous Galerkin method is unconditionally well-posed

- The average $\{v\}_S$ is well defined if $v \in H^{1/2}(\mathcal{T}_h)$ and the jump $[v]_S$ is well defined if $v \in H^{1/2-\gamma}(\mathcal{T}_h)$

We have $S^1(\mathcal{T}_h) \subset S^1,dG(\mathcal{T}_h)$ and $S^{1,cr}(\mathcal{T}_h) \subset S^{1,dG}(\mathcal{T}_h)$

**A.7.5 Projects**

**Project A.7.1** Implement the pressure-stabilized $P1-P1$ method for the Stokes problem and investigate the experimental convergence using the model solution

$$u(x, y) = \pi \begin{bmatrix} \sin(2\pi y) \sin^2(\pi x) \\ -\sin(2\pi x) \sin^2(\pi y) \end{bmatrix},$$

$$p(x, y) = \cos(\pi x) \sin(\pi y)$$
in the square \( \Omega = (-1, 1)^2 \) with homogeneous Dirichlet boundary conditions on \( \Gamma_D = \partial \Omega \) for \( u \).

**Project A.7.2** Determine experimental convergence rates for approximating the Stokes problem with the nonconforming Crouzeix–Raviart method for the model solution

\[
\begin{align*}
    u(x, y) &= \pi \begin{bmatrix} \sin(2\pi y) \sin^2(\pi x) \\ -\sin(2\pi x) \sin^2(\pi y) \end{bmatrix}, \\
p(x, y) &= \cos(\pi x) \sin(\pi y)
\end{align*}
\]

in the square \( \Omega = (-1, 1)^2 \) with homogeneous Dirichlet boundary conditions on \( \Gamma_D = \partial \Omega \) for \( u \).

**Project A.7.3** Consider the dual mixed formulation of the Poisson problem and its approximation using the Raviart–Thomas finite element method. Determine experimental convergence rates for both variables on sequences of uniformly refined triangulations for the domain \( \Omega = (0, 1)^2 \) with exact solution \( u(x, y) = \sin(\pi x) \sin(\pi y) \), and the domain \( \Omega = (-1, 1)^2 \setminus ([-1, 0] \times [0, 1]) \) with the exact solution \( u(r, \phi) = r^{2/3} \sin(2\phi/3) \) in polar coordinates.

**Project A.7.4** Implement the \( P_1-P_0 \) finite element method for the dual mixed formulation of the Poisson problem and demonstrate experimentally that it is ill-posed in general. Try to stabilize the method by incorporating an appropriate penalty term.

**Project A.7.5** Determine a function \( f \in L^2(\Omega; \mathbb{R}^2) \) with \( \Omega = (-1, 1)^2 \) so that the exact solution \( u \in H^1_0(\Omega; \mathbb{R}^2) \) of the variational formulation

\[
\int_\Omega \nabla u : \nabla v \, dx + \varepsilon^{-2} \int_\Omega \text{div} \, u \, \text{div} \, v \, dx = \int_\Omega f \cdot v \, dx
\]

for all \( v \in H^1_0(\Omega; \mathbb{R}^2) \) is given by

\[
u(x_1, x_2) = \begin{bmatrix} \sin(2\pi x_2) \sin^2(\pi x_1) \\ -\sin(2\pi x_1) \sin^2(\pi x_2) \end{bmatrix}.
\]

Introduce the variable \( p = \varepsilon^{-2} \text{div} \, u \), and rewrite the problem as a saddle-point formulation with penalty term. Discretize it with a nonconforming method and investigate the experimental convergence of approximations for the parameters \( \varepsilon = 10^{-j}, j = 1, 2 \).
Project A.7.6 A simple mathematical description of the release of smoke from a chimney and its distribution in the environment leads to the convection dominated equation

$$-\nu \Delta c + b \cdot \nabla c = 0$$

for the smoke concentration $c$, the diffusion coefficient $\nu = 1.5 \cdot 10^{-5} \text{m}^2/\text{s}$ of carbon dioxide in air, and the velocity field $b = [10, 0, 0]^T \text{m/s}$. Assume that the chimney is 50 m high, and simulate a two-dimensional model reduction of the problem with appropriate boundary conditions up to a height of 200 m and a distance of 1000 m in the direction of the wind. Compare a direct approximation with a stabilized one.

Project A.7.7 Experimentally determine the experimental convergence rate of different discontinuous Galerkin methods for the Poisson problem $-\Delta u = f$ in $\Omega = (0, 1)^2$ with boundary condition $u|_{\partial \Omega} = 0$ on a sequence of uniformly refined triangulations for the exact solution

$$u(x, y) = \sin(\pi x) \sin(\pi y).$$

Investigate also dependence on the parameter $\gamma$.

Project A.7.8 Incorporate the treatment of convective terms in implementing the discontinuous Galerkin method and test its performance with meaningful experiments.

A.8 Applications

A.8.1 Linear Elasticity

Exercise A.8.1

(i) Show that $SO(d) = \{Q \in \mathbb{R}^{d \times d} : Q^T Q = I, \det Q = 1\}$ is a $d(d - 1)/2$-dimensional submanifold in $\mathbb{R}^{d \times d}$.

(ii) Prove that the tangent space of $SO(d)$ at the identity matrix is given by $T_I SO(d) = so(d) = \{U \in \mathbb{R}^{d \times d} : U^T + U = 0\}$.

Exercise A.8.2 Assume that $v \in H^1(\Omega; \mathbb{R}^d)$ for $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ satisfies $\nabla v(x) \in SO(d)$ for almost every $x \in \Omega$.

(i) Show that $\text{div} \text{Cof} \nabla w = 0$ for $w \in H^1(\Omega; \mathbb{R}^d)$, where $\text{Cof} A = (\det A)A^{-T}$, and conclude that every component of $v$ is harmonic.

(ii) Show that $\Delta v = 0$ implies $\Delta (|\nabla v|^2 - d) = 2|D^2 v|^2$ and deduce that $\nabla v$ is constant.
Exercise A.8.3  Construct a solution \( u(x) = Ax \) with a diagonal matrix \( A \in \mathbb{R}^{3 \times 3} \) of the Navier–Lamé equations in the cylinder domain \((-L/2, L/2) \times B_r(0) \subset \mathbb{R}^3 \) with \( \Gamma_D = \emptyset, \Gamma_N = \partial \Omega \), and

\[
g(x) = \begin{cases} \pm e_1, & x_1 = \pm L/2, \\ 0, & -L/2 < x_1 < L/2, \end{cases}
\]

for \( x = (x_1, x_2, x_3) \in \Gamma_N \). Determine the ratio between elongation and radial compression and sketch the solution for different Lamé constants \( (\lambda, \mu) \).

Exercise A.8.4  For \( \lambda, \mu > 0 \) let \( \mathbb{C} : \mathbb{R}^{d \times d}_{\text{sym}} \to \mathbb{R}^{d \times d}_{\text{sym}} \) be defined by \( \mathbb{C}A = \lambda \, \text{tr}(A)I + 2\mu A \). Show that \( \mathbb{C} \) is invertible with

\[
\mathbb{C}^{-1}B = \frac{1}{2\mu} \left( B - \frac{\lambda}{d\lambda + 2\mu} \, \text{tr}(B)I \right).
\]

Exercise A.8.5  We consider the elastic deformation of a solid occupying the domain \( \Omega \subset \mathbb{R}^3 \).

(i) Assume that the expected behavior of the solid is such that for \( \varepsilon = \varepsilon(u) \in \mathbb{R}^{3 \times 3} \), we have \( \varepsilon_{ij} = 0 \) for \( i = 1, 2, 3 \). Derive a simplified two-dimensional model.

(ii) Assume that the expected behavior of the solid is such that for \( \sigma = \mathbb{C}\varepsilon(u) \in \mathbb{R}^{3 \times 3} \), we have \( \sigma_{ij} = 0 \) for \( i = 1, 2, 3 \). Derive a simplified two-dimensional model.

(iii) Discuss model situations for which the simplifications apply.

Exercise A.8.6  Assume that \( u \in H^1_D(\Omega; \mathbb{R}^d) \) is minimal for

\[
I(u) = \frac{1}{2} \int_{\Omega} \mathbb{C}\varepsilon(u) : \varepsilon(u) \, dx - \int_{\Omega} f \cdot u \, dx - \int_{\Gamma_N} g \cdot u \, ds.
\]

Prove that \( u \) is a weak solution of the Navier–Lamé equations.

Exercise A.8.7

(i) Prove that for \( v \in C^2(\overline{\Omega}; \mathbb{R}^d) \) we have

\[
2 \, \text{div} \, \varepsilon(v) = \Delta v + \nabla \, \text{div} \, v.
\]
(ii) Assume that $v(x) = Ax + b$ for $x \in \Omega$ with a skew-symmetric matrix $A \in \mathbb{R}^{d \times d}$ such that $v|_{\Gamma_D} = 0$. Show that $v = 0$.

**Exercise A.8.8** Devise and analyze a numerical method for approximating the problem of determining $(\sigma, u) \in L^2(\Omega; \mathbb{R}^{d \times d}_{\text{sym}}) \times H^1_D(\Omega; \mathbb{R}^d)$ such that

$$
\int_\Omega C^{-1}\sigma : \tau \, dx - \int_\Omega \tau : \varepsilon(u) \, dx = 0,
$$

$$
\int_\Omega \sigma : \varepsilon(v) \, dx = \ell(v),
$$

for all $(\tau, v) \in L^2(\Omega; \mathbb{R}^{d \times d}_{\text{sym}}) \times H^1_D(\Omega; \mathbb{R}^d)$. Discuss the dependence of the approximation error on the Lamé constants.

**Exercise A.8.9** Consider the Navier–Lamé equations with $\Gamma_D = \partial \Omega$, and let $u_h \in \mathcal{H}^{1,cr}(\mathcal{T}_h)^d$ satisfy

$$
\mu \int_\Omega \nabla_T u_h : \nabla_T v_h \, dx + (\mu + \lambda) \int_\Omega \text{div}_T u_h \text{div}_T v_h \, dx = \int_\Omega f \cdot v_h \, dx
$$

for all $v_h \in \mathcal{H}^{1,cr}(\mathcal{T}_h)^d$. Show that the approximations converge to the solution of the Navier–Lamé equations as $h \to 0$.

**Exercise A.8.10** Assume that a discrete Korn inequality holds on $\mathcal{H}^{1,cr}(\mathcal{T}_h)^d$.

(i) Show that there exists a unique $u_h \in \mathcal{H}^{1,cr}(\mathcal{T}_h)^d$ with

$$
2\mu \int_\Omega \varepsilon_T(u_h) : \varepsilon_T(v_h) \, dx + \lambda \int_\Omega \text{div}_T u_h \text{div}_T v_h \, dx = \ell(v_h)
$$

for all $v_h \in \mathcal{H}^{1,cr}(\mathcal{T}_h)^d$.

(ii) Assume that the exact solution of the Navier–Lamé equations satisfies

$$
\|u\|_{H^2(\Omega)} + \lambda \|\nabla \text{div} u\|_{L^2(\Omega)} \leq \text{c}_{NL}\|f\|_{L^2(\Omega)}.
$$

Prove that

$$
\|C^{1/2}\varepsilon_T(u - u_h)\|_{L^2(\Omega)} \leq \text{c}_{cr}(2\mu)^{1/2} + \lambda^{-1/2})h\|f\|_{L^2(\Omega)}.
$$
**Quiz A.8.1** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>A discrete Korn inequality is used in deriving the error estimate $|C^{1/2}\nabla_y(u - u_b)|<em>{L^2(\Omega)} \leq c h |D^2u|</em>{L^2(\Omega)}$</td>
<td></td>
</tr>
<tr>
<td>The sets $\mathbb{R}^{d\times d}_{\text{sym}}$ and $so(d)$ are orthogonal with respect to the scalar product $A : B = \text{tr}A^TB$</td>
<td></td>
</tr>
<tr>
<td>For all matrices $A, B \in \mathbb{R}^{d\times d}$ we have $A : B^T = A^T : B = B^T : A$</td>
<td></td>
</tr>
<tr>
<td>If the stresses $\sigma = C\varepsilon(u)$ vanish, then the displacement $u$ is a linearized rigid body motion</td>
<td></td>
</tr>
<tr>
<td>Crouzeix–Raviart finite elements avoid incompressibility locking but may lead to ill-posed discrete problems</td>
<td></td>
</tr>
</tbody>
</table>

### A.8.2 Plate Bending

**Exercise A.8.11** Assume that the Poisson problem is $H^2$-regular in $\omega$, and let $x_0 \in \overline{\omega}$. Show that there exists a unique weak solution $u \in H^2(\omega)$ of the boundary value problem

$$\Delta^2 u = \delta_{x_0} \text{ in } \omega, \quad u = \nabla u = 0 \text{ on } \partial\omega,$$

where $\delta_{x_0} : H^2(\omega) \to \mathbb{R}$ is defined by $\delta_{x_0}(v) = v(x_0)$ for all $v \in H^2(\omega)$.

**Exercise A.8.12**

(i) Prove that $V = \{v \in H^2(\omega) : v = \nabla v = 0 \text{ on } \partial\omega\}$ is a closed subspace of $H^2(\omega)$.

(ii) Show that $V$ coincides with the closure of $C_0^\infty(\omega)$ with respect to the Sobolev norm $\| \cdot \|_{H^2(\omega)}$.

**Exercise A.8.13** Show that $u \in H^2(\omega) \cap H_0^1(\omega)$ is a minimizer for the Kirchhoff functional

$$I_{Ki}(u) = \frac{\theta}{2} \int_\omega |\Delta u|^2 \, dx + \frac{1 - \theta}{2} \int_\omega |D^2u|^2 \, dx - \int_\omega fu \, dx$$

if and only if $a(u, v) = \ell(v)$ for all $v \in H^2(\omega) \cap H_0^1(\omega)$ with

$$a(u, v) = \int_\omega \Delta u \Delta v \, dx + (1 - \theta) \int_\omega 2 \partial_1 \partial_2 u \partial_1 \partial_2 v - \partial_1^2 \partial_1^2 v - \partial_1^2 v \partial_2^2 u \, dx,$$

$$\ell(v) = \int_\omega fv \, dx.$$
Exercise A.8.14  Show that for $u \in H^4(\omega)$ and $v \in H^2(\omega)$, we have
\[
\int_\omega \Delta u \Delta v \, dx = \int_\omega \Delta^2 u v \, dx - \int_{\partial \omega} (\partial_n \Delta u) v \, ds + \int_{\partial \omega} \Delta u \partial_n v \, ds.
\]

Exercise A.8.15 Let $\phi \in L^2(\partial \omega)$ and assume that
\[
\int_{\partial \omega} \phi \partial_n v \, ds = 0
\]
for all $v \in H^2(\omega)$. Show that $\phi = 0$ on $\partial \omega$. Assume first that for $x_0 \in \partial \omega$, there exists $\varepsilon > 0$ such that $B_\varepsilon(x_0) \cap \partial \omega \subset \mathbb{R} \times \{0\}$.

Exercise A.8.16 Let $f = 1$ in $\omega = B_1(0) \subset \mathbb{R}^2$, and $\theta \in (0, 1)$. Determine the solutions of
\[
\Delta^2 u = f \text{ in } \omega, \quad u = \Delta u = 0 \text{ on } \partial \omega,
\]
and
\[
\Delta^2 u = f \text{ in } \omega, \quad u = \Delta u + (1 - \theta) \kappa \partial_n u = 0 \text{ on } \partial \omega.
\]
Discuss their difference and the approximation of the problems on polygonal domains $\omega_h \subset \omega$.

Exercise A.8.17 Consider the boundary value problem
\[
\Delta^2 u = f \text{ in } \omega, \quad u = \Delta u = 0 \text{ on } \partial \omega.
\]
Introduce the variable $v = \Delta u$ and formulate and analyze an equivalent saddle-point formulation under appropriate conditions on $\omega$. Discuss error estimates for the approximation of the saddle-point formulation with a low order finite element method.

Exercise A.8.18 Let $u_D = \tilde{u}_D|_{\partial \omega}$ for some $\tilde{u}_D \in H^2(\omega)$ and consider the minimization of the functionals
\[
I_1(u) = \frac{1}{2} \int_\omega |\nabla u|^2 \, dx, \quad I_2(u) = \frac{1}{2} \int_\omega |\Delta u|^2 \, dx,
\]
subject to $u|_{\partial \omega} = u_D$.

(i) Show that the problems define under appropriate assumptions on $u_D$ surfaces of minimal area and minimal total curvature.

(ii) Derive the Euler–Lagrange equations for both minimization problems.

(iii) Compute the solutions for $\omega = B_1(0)$ and $u_D(\theta) = \sin(\theta)$, $\theta \in [0, 2\pi]$. 

Exercise A.8.19  Let $a : H^2(\omega) \times H^2(\omega) \to \mathbb{R}$ be the bilinear form associated with the Kirchhoff bending energy, and let $\ell \in H^2(\omega)'$.

(i) Show that there exists a unique function $u_h \in S_{h_{0,1}}(\mathcal{T}_h)$ such that

$$a(u_h, v_h) = \ell(v_h)$$

for all $v_h \in S_{h_{0,1}}(\mathcal{T}_h)$.

(ii) Assume that the exact solution of the plate bending problem satisfies $u \in H^4(\omega)$, and prove that we have

$$\|u - u_h\|_{H^2(\omega)} \leq c_{\text{Arg}} h^2 \|u\|_{H^4(\omega)}.$$

Exercise A.8.20  Let $T \subset \mathbb{R}^2$ be a triangle with vertices $z_0, z_1, z_2$, midpoint $x_T$ and sides $S_0, S_1, S_2$. Let $K_0, K_1, K_2$ be the subtriangles with vertex $x_T$ and sides $S_0, S_1, S_2$, respectively. For $i = 0, 1, 2$ and $\alpha \in \mathbb{N}_0^2$ with $|\alpha| \leq 1$, and $v \in C^1(T)$, define

$$\chi_{i,\alpha}(v) = \partial^\alpha v(z_i), \quad \chi_{i,n}(v) = \nabla v(x_{S_i}) \cdot n_{S_i},$$

with midpoints $x_{S_i}$ and normals $n_{S_i}$ for the sides $S_i$, $i = 0, 1, 2$. Show that if $v \in C^1(T)$ with $v|_{K_i} \in P_3(K_i)$, $i = 0, 1, 2$, and

$$\chi_{i,\alpha}(v) = 0, \quad \chi_{i,n}(v) = 0$$

for $i = 0, 1, 2$, and $\alpha \in \mathbb{N}_0^2$ with $|\alpha| \leq 1$, then we have $v = 0$.

Quiz A.8.2  Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

- Plate bending refers to a model reduction of linear elasticity corresponding to small thickness $t > 0$
- The bending problem with simple support boundary conditions is unconditionally well-posed
- We have that $S_{h_{0,1}}(\mathcal{T}_h) \subset H^2(\omega) \cap H^1_0(\omega)$
- If a piecewise polynomial function belongs to $H^2(\omega)$, then its derivatives are continuous
- The Argyris element leads to quadratic convergence rates if $u \in H^4(\omega)$
A.8.3 Electromagnetism

Exercise A.8.21  Let $w, \phi \in C^1(\Omega; \mathbb{R}^3)$.

(i) Show that we have
\[
\text{div}(\phi \times w) = w \cdot \text{curl} \phi - \phi \cdot \text{curl} w.
\]

(ii) Prove that
\[
\int_{\Omega} w \cdot \text{curl} \phi \, dx = \int_{\Omega} \text{curl} w \cdot \phi \, dx - \int_{\partial \Omega} (w \wedge n) \cdot \phi \, ds.
\]

Exercise A.8.22

(i) Prove that $H_0(\text{curl}; \Omega)$ is a Banach space.

(ii) Show that $\nabla H_0^1(\Omega) \subset H_0(\text{curl}; \Omega)$.

(iii) Prove that for all $v \in H(\text{curl}; \Omega)$ we have $\text{div} \, \text{curl} \, v = 0$.

Exercise A.8.23

(i) Show that for $\Omega \subset \mathbb{R}^d$ and $\phi \in C^2(\Omega; \mathbb{R}^d)$, we have
\[
\text{Curl} \, \text{curl} \, \phi = \nabla \text{div} \phi - \Delta \phi,
\]

and if $d = 2$, $\text{curl} \, \text{Curl} \psi = \Delta \psi$ for every $\psi \in C^2(\Omega)$.

(ii) Show that in the absence of charges and currents, solutions $E$ and $B$ of the Maxwell system in free space are solutions of wave equations with wave speed that coincide with the speed of light. Use that $\varepsilon_0 = 8.854 \, 187 \cdot 10^{-12} \text{F/m}$ and $\mu_0 = 4\pi \cdot 10^{-7} \text{N/A}^2$.

Exercise A.8.24  Let $F \in H(\text{div}; \Omega)$ with $F \cdot n = 0$ on $\partial \Omega$. Show that there exist functions $\phi \in H^1(\Omega)$ and $G \in H(\text{div}; \Omega)$ with $G \cdot n = 0$ on $\partial \Omega$ and $\text{div} \, G = 0$, such that
\[
F = \nabla \phi + G.
\]

Show that the decomposition is $L^2$-orthogonal.

Exercise A.8.25  Let $G \in C^1(\mathbb{R}^3; \mathbb{R}^3)$. Show that there exists $\phi \in C^1(\mathbb{R}^3)$ such that
\[
G = \nabla \phi
\]
if and only if $\text{curl} \, G = 0$. 
Exercise A.8.26

(i) Show that the solution operator \((-\Delta)^{-1} : L^2(\Omega) \rightarrow L^2(\Omega)\) related to the Poisson problem with homogeneous Dirichlet boundary conditions is bounded, compact, and self-adjoint.

(ii) Show that the boundary value problem

\[-\Delta u - \omega^2 u = f \text{ in } \Omega, \quad u|_{\partial\Omega} = 0.\]

admits for every \(f \in L^2(\Omega)\) a unique weak solution, provided that \(\omega^2\) does not coincide with an eigenvalue of the operator \(-\Delta\).

Exercise A.8.27 Let \(\Omega \subset \mathbb{R}^2\). Show that every eigenfunction of the Laplace operator with homogeneous Neumann boundary conditions defines a solution of the Maxwell eigenvalue problem with constraint \(\text{div } u = 0\).

Exercise A.8.28 Let \(v_h \in C(\overline{\Omega}; \mathbb{R}^d)\) be a piecewise polynomial vector field. Show that \(v_h \in H_0(\text{curl}; \Omega)\) if and only if the tangential component of \(v_h\) is continuous and vanishes on the boundary.

Exercise A.8.29

(i) Prove that we have \(\nabla \mathcal{I}_0^1(\mathcal{T}_h) \subset \mathcal{N} \cup \mathcal{D}_h(\mathcal{T}_h)\).

(ii) Show that \(\mathcal{I}_h \nabla \phi = \nabla \mathcal{I}_h \phi\) for all \(\phi \in C^1(\overline{\Omega})\).

Exercise A.8.30

(i) Let \(v \in C^1(\mathbb{R}^3; \mathbb{R}^3)\) with \(\text{div } v = 0\). Assume that there exists \(\psi \in C^1(\mathbb{R}^3; \mathbb{R}^3)\) such that \(v = \text{curl } \psi\). Show that there exists \(\tilde{\psi} \in C^1(\mathbb{R}^3; \mathbb{R}^3)\) with \(\tilde{\psi}_3 = 0\) and \(v = \text{curl } \tilde{\psi}\).

(ii) Let \(v(x, y, z) = [x^2, 3xz^2, -2xz]^T\). Construct \(\psi \in C^1(\mathbb{R}^3; \mathbb{R}^3)\) such that \(v = \text{curl } \psi\).

Quiz A.8.3 Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Statement</th>
<th>True/False</th>
</tr>
</thead>
<tbody>
<tr>
<td>For (d = 2) we have (v \land n = \det[v, n]) and for (d = 3), we have (v \cdot (a \times b) = \det[v, a, b])</td>
<td></td>
</tr>
<tr>
<td>If (a \in \mathbb{R}^3) and (a \times e_i = 0) for (i = 1, 2, 3), then (a = 0)</td>
<td></td>
</tr>
<tr>
<td>If (v \land n = 0) on (\partial\Omega), then (v) is parallel to (n) on (\partial\Omega)</td>
<td></td>
</tr>
<tr>
<td>We have (\text{curl } v_h \in \mathcal{I}_h^1(\mathcal{T}_h)) for every (v_h \in \mathcal{R}^0(\mathcal{T}_h)).</td>
<td></td>
</tr>
<tr>
<td>We have that (\nabla \phi_z) is orthogonal to every side (S \in \mathcal{T}_h) with (z \notin S).</td>
<td></td>
</tr>
</tbody>
</table>
\textbf{A.8.4 Incompressible, Viscous Fluids}

\textbf{Exercise A.8.31} Let $A \in C^1([0, T]; \mathbb{R}^{d \times d})$ and $t_0 \in [0, T]$ be such that $A(t_0)$ is invertible. Show that we have

$$\frac{d}{dt} \left|_{t=t_0} \right. \det A(t) = \det A(t_0) \text{tr} \left( A(t_0)^{-1} A'(t_0) \right), \quad A'(t) = \frac{d}{dt} A(t)$$

\textit{Hint:} Use Leibniz’s formula to prove the identity for the special case $A(t_0) = I$ first.

\textbf{Exercise A.8.32} Let $\Phi : [0, T] \times \Omega \rightarrow \mathbb{R}^d$ be a $C^2$ mapping, such that $\Phi(t, \cdot)$ defines a diffeomorphism between $\Omega$ and $\Omega_t = \Phi(t, \Omega)$ for every $t \in [0, T]$. Assume that $\Phi(0, x) = x$ for all $x \in \Omega$, i.e., $\Omega_0 = \Omega$, and that $J(t, x) = \det D\Phi(t, x) > 0$ for all $(t, x) \in [0, T] \times \Omega$.

(i) For $t \in [0, T]$, $x \in \Omega$, and $y = \Phi(t, x)$ let $v(t, y) = \partial_t \Phi(t, x)$. Show that

$$\partial_t J(t, x) = J(t, x) \text{div} v((t, \Phi(t, x))).$$

(ii) For $t \in [0, T]$ and $y \in \Omega_t$, let $\varrho(t, y)$ be the mass density of a material occupying the domain $\Omega_t$. The mass of the set $\omega_t = \Phi(t, \omega)$ for $\omega \subset \Omega$ is given by

$$m_{\omega_t} = \int_{\omega_t} \varrho(t, y) \, dy.$$

Assume that mass is conserved to deduce that

$$\partial_t \rho + \text{div}(\rho v) = 0.$$

What can you conclude for incompressible materials, i.e., when $t \mapsto m_{\omega_t}$ is constant for every $\omega \subset \Omega$?

\textbf{Exercise A.8.33} For a vector field $u : \Omega \rightarrow \mathbb{R}^d$, let

$$\sigma = 2\mu \varepsilon(u) + \lambda \text{tr}(u) I - pI.$$

Show that

$$\text{div} \sigma = \mu \Delta u + (\lambda + \mu) \nabla \cdot u - \nabla p,$$

where the divergence is taken row-wise.
Exercise A.8.34 Prove that the trilinear form \( n : [H^1_0(\Omega; \mathbb{R}^d)]^3 \to \mathbb{R} \),

\[
n(z; u, v) = \int_\Omega (z \cdot \nabla u) \cdot v \, dx,
\]
is skew-symmetric in the second and third variable, provided that \( \text{div} \, z = 0 \).

Exercise A.8.35 Show that a non-dimensionalization of the stationary Navier–Stokes equations with a characteristic length \( L \), e.g., \( L = \text{diam}(\Omega) \), a characteristic speed \( U \), e.g., \( U = \max_{x \in \Omega} |u_D(x)| \), leads to the system of equations

\[
- \frac{1}{R} \Delta \tilde{u} + \tilde{u} \cdot \nabla \tilde{u} + \nabla \tilde{p} = \frac{L}{U^2} \tilde{f}, \quad \text{div} \, \tilde{u} = 0,
\]

where \( R = UL/\nu \).

Exercise A.8.36 Let \( \tilde{u} \in H_N(\text{div}; \Omega) \) and \((u, p) \in H_N(\text{div}; \Omega) \times L^2_0(\Omega)\) be a weak solution of the system

\[
u + \nabla p = \tilde{u}, \quad \text{div} \, u = 0.
\]

Show that \( u \) is the \( L^2 \)-projection of \( \tilde{u} \) onto the space \( \{ v \in H_N(\text{div}; \Omega) : \text{div} \, v = 0 \} \).

Exercise A.8.37 Show that if \( \Gamma_D = \partial \Omega \) and \( ||f||_{L^2(\Omega)} \leq v^2/(c_P c_S^2) \), where \( c_P, c_S > 0 \) are such that \( ||v||_{L^2(\Omega)} \leq c_P ||\nabla v||_{L^2(\Omega)} \) and \( ||v||_{L^4(\Omega)} \leq c_S ||\nabla v||_{L^2(\Omega)} \) for all \( v \in H^1_0(\Omega; \mathbb{R}^d) \), solutions of the stationary Navier–Stokes equations are unique.

Exercise A.8.38

(i) Let \( 1 \leq p, q, r \leq \infty \) and \( u \in L^p(\Omega), v \in L^q(\Omega), \) and \( w \in L^r(\Omega) \). Show that

\[
\int_\Omega u v w \, dx \leq ||u||_{L^p(\Omega)} ||v||_{L^q(\Omega)} ||w||_{L^r(\Omega)}
\]

provided that \( 1/p + 1/q + 1/r = 1 \).

(ii) For which exponents \( 1 \leq p, q, r \leq \infty \) and dimensions \( 1 \leq d \leq 3 \) is the trilinear form

\[
n : W^{3,p}(\Omega) \times W^{1,q}(\Omega) \times L^r(\Omega) \to \mathbb{R}, \quad n(u, v, w) = \int_\Omega (\Delta u) v w \, dx,
\]

bounded?

Exercise A.8.39 Brouwer’s fixed point theorem states that every continuous mapping \( f : C \to C \) on a nonempty, convex, and compact set \( C \subset \mathbb{R}^n \) has a fixed point.
Prove via contradiction that for every continuous mapping \( F : \mathbb{R}^n \to \mathbb{R}^n \) with the property that
\[
F(U) \cdot U \geq 0
\]
for all \( U \in \mathbb{R}^n \) with \(|U| \geq R > 0\), there exists \( U^* \in \mathbb{R}^n \) with \(|U^*| \leq R\) such that \( F(U^*) = 0 \).

**Exercise A.8.40**

(i) Use the closed range theorem to show that if \( \phi \in H_0^1(\Omega; \mathbb{R}^d)' \) is such that
\[
\phi(v) = 0
\]
for all \( v \in H_0^1(\Omega; \mathbb{R}^d) \) with \( \text{div} \ v = 0 \), then there exists \( p \in L_0^2(\Omega) \) such that
\[
\phi(v) = \int_\Omega p \ \text{div} \ v \ \text{d}x
\]
for all \( v \in H_0^1(\Omega; \mathbb{R}^d) \).

(ii) Conclude that it suffices to determine \( u \in H_0^1(\Omega; \mathbb{R}^d) \) as the solution of the equation
\[
v \int_\Omega \nabla u : \nabla v \ \text{d}x + \int_\Omega (u \cdot \nabla u) \cdot v \ \text{d}x = \int_\Omega f \cdot v \ \text{d}x
\]
subject to \( \text{div} \ u = 0 \) and for all \( v \in H_0^1(\Omega; \mathbb{R}^d) \) with \( \text{div} \ v = 0 \), in order to solve the stationary Navier–Stokes equations.

**Quiz A.8.4** Decide for each of the following statements whether it is true or false. You should be able to justify your decision.

<table>
<thead>
<tr>
<th>Solutions of the stationary Navier–Stokes equations are unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>If ( \Gamma_D = \partial \Omega ) and ( u_D = 0 ), then the Picard iteration is globally convergent</td>
</tr>
<tr>
<td>The vector field ( u : \Omega \to \mathbb{R}^d ) in the Navier–Stokes equations determines the displacements of particles in the domain ( \Omega )</td>
</tr>
<tr>
<td>If ( a : V \times V \to \mathbb{R} ) is coercive and ( n : V \times V \to \mathbb{R} ) skew-symmetric, then ( a + \lambda n ) is coercive on ( V \times V ) for every ( \lambda \in \mathbb{R} )</td>
</tr>
<tr>
<td>The Stokes system is a linearization of the Navier–Stokes equations</td>
</tr>
</tbody>
</table>
A.8.5 Projects

Project A.8.1 Implement the discretization of the Navier–Lamé equations with a $P1$ finite element method and test your code using the exact solution

$$u(t, x, y) = \begin{bmatrix} \sin(2\pi y) \sin^2(\pi x) \\ -\sin(2\pi x) \sin^2(\pi y) \end{bmatrix},$$

by defining the right-hand side $f$ in $\Omega = (0, 1)^2$ appropriately. Illustrate the failure of the method in the case of a nearly incompressible material, i.e., investigating the dependence of the approximation error on $\lambda = 10^j, j = 1, 2, \ldots, 6$. Compare this to the approximation with the stabilized Crouzeix–Raviart method.

Project A.8.2 Consider a cylinder domain $\Omega$ of length $\ell = 1$ and radius $r = 1/10$, i.e., $\Omega = B_r(0) \times (0, \ell) \subset \mathbb{R}^3$ that represents an elastic rod. We assume that the rod is fixed on the side $\Gamma_0 = B_r(0) \times \{0\}$. Undamped vibrations of the rod are then described by the equation

$$\partial_t^2 u - \text{div} \mathbf{\varepsilon}(u) = 0 \quad \text{in} \ (0, T) \times \Omega$$

supplemented with boundary and initial conditions. Devise a weak formulation and a numerical method for simulating the vibrations and carry out experiments with different discretization parameters.

Project A.8.3 Use the Argyris element to discretize the Poisson problem with homogeneous Dirichlet boundary conditions on the unit square. Determine the experimental convergence rate for the exact solution

$$u(x, y) = \sin(\pi x) \sin(\pi y).$$

Improve the conditioning of the local linear systems of equations by using scaled monomials $p_{jk}(x_1, x_2) = h_T^{-(j+k)}(x_1 - x_{T,1})^j(x_2 - x_{T,2})^k$.

Project A.8.4 We consider a simply supported plate occupying the domain $\Omega = B_1(0)$ with force $f = 1$ and material parameter $\theta = 1/2$, i.e., the boundary value problem

$$\Delta^2 u = 1 \quad \text{in} \ \Omega,$$

$$u = 0 \quad \text{on} \ \partial \Omega,$$

$$\Delta u - \frac{1}{2} \partial_n u = 0 \quad \text{on} \ \partial \Omega.$$

Verify that for a rotationally symmetric function, we have $\Delta u = \partial_r^2 u + r^{-1}\partial_r u$, $\partial_n u = \partial_r u$, and $\Delta^2 u = (\partial_r^4 + 2r^{-1}\partial_r^3 - r^{-2}\partial_r^2 + r^{-3}\partial_r) u$, so that the exact solution
is given by
\[ u(r) = \frac{1}{64} r^4 - \frac{14}{3 \cdot 64} r^2 + \frac{11}{3 \cdot 64}. \]

Approximate the problem on a sequence of polygonal domains \((\Omega_n)_{n=0,1,\ldots}\) and show experimentally that the approximations \((u_n)_{n=0,1,\ldots}\) do not converge to \(u\) by comparing \(u_n(0)\) for \(n = 0, 1, \ldots\) with \(u(0)\). Show that in the case of clamped boundary conditions, the exact solution is given by \(u(r) = r^4/64 - r^2/32 + 1/64\) and that the problem does not occur.

**Project A.8.5** Implement the approximation of the Maxwell equations with a \(P1\)-finite element method and test its performance on the unit square \(\Omega = (0,1)^2\) with \(f = [1,1]^T\), the \(L\)-shaped domain \(\Omega = (-1,1)^2 \setminus (-1,0) \times (0,1)\) with \(f = [1,1]^T\), and the ring domain \(\Omega = B_1(0) \setminus B_{1/2}(0)\) with \(f(x) = \text{Curl} \ l(x)\). Compare approximations qualitatively for mesh-sizes \(h \approx 1/20\) to approximations obtained with the Nédélec method for different choices of \(\omega^2\).

**Project A.8.6** Verify that \(\lambda = \pi^2\) is an eigenvalue of the Maxwell operator in the unit square \(\Omega = (0,1)^2\). For a fixed triangulation of \(\Omega\) with \(h \approx 1/100\), investigate the dependence of the approximation error on the difference \(|\omega^2 - \lambda|\), by considering \(\omega^2 = \lambda + 1/10^j\) for \(j = 0, 1, 2, \ldots\) and constructing an appropriate reference solution.

**Project A.8.7** Use the Crouzeiz–Raviart method and a backward difference quotient to discretize the Stokes flow
\[ \partial_t u - \Delta u + \nabla p = f, \quad \text{div } u = 0, \]
in \((0, T) \times \Omega\) with no-slip boundary conditions for \(u\) on \(\partial \Omega\). Define \(f : (0, T) \times \Omega \to \mathbb{R}^2\) so that the exact solution of the problem with \(\Omega = (-1,1)^2\) and \(T = 1\) is given by
\[ u(t,x,y) = \pi \sin(t) \begin{bmatrix} \sin(2\pi y) \sin^2(\pi x) \\ -\sin(2\pi x) \sin^2(\pi y) \end{bmatrix}, \]
\[ p(t,x,y) = \sin(t) \cos(\pi x) \sin(\pi y). \]
Use a fixed mesh-size \(0 < h \ll 1\) and different step-sizes \(\tau > 0\) to investigate the convergence behavior of the errors
\[ \max_{k=0,\ldots,K} \|p - p_h^k\|_{L^2(\Omega)}, \quad \max_{k=0,\ldots,K} \|\nabla \varphi(u - u_h^k)\|_{L^2(\Omega)} \]
as \(\tau \to 0\). Compare the convergence behavior to that for approximations obtained with the Chorin projection scheme. Plot the pressure errors at the final time \(T\) and discuss your observations.
Project A.8.8 Use the Picard iteration and a Crouzeiz–Raviart discretization to solve the stationary Navier–Stokes equations in the cylinder domain with hole \( \Omega = \{(-\ell, \ell) \times B_r(0) \} \setminus B_{r/2}(0) \subseteq \mathbb{R}^d \) for \( \ell = 3 \), \( r = 1 \), and \( d = 2, 3 \), for the Dirichlet boundary conditions

\[
u(x_1, x_2, \ldots, x_d) = \begin{cases} 
  x_2^2 + \cdots + x_d^2 - 1, & x_1 = \pm \ell, \\
  0, & -\ell < x_1 < \ell.
\end{cases}
\]

Use a mesh-size \( h \approx 1/20 \) and test the convergence behavior of the iteration for relative viscosities \( \nu = 10^{-j}, j = 0, 1, \ldots, 3 \). Replace the Dirichlet boundary condition at \( x_1 = \ell \) by a homogeneous Neumann boundary condition and repeat the experiment. Visualize some solutions with paraview.
Appendix B
Implementation Aspects

B.1 Basic MATLAB Commands

B.1.1 Matrix Operations

The programming language MATLAB provides various optimized implementations of matrix operations. Some important commands, whose usage is canonical, are listed in Table B.1. Various operations such as matrix multiplication can be applied component-wise by placing a dot in front of the operand, e.g.,

\[
\begin{bmatrix}
1 & 2 \\
3 & 4
\end{bmatrix}
\cdot
\begin{bmatrix}
5 & 6 \\
7 & 8
\end{bmatrix}
= \begin{bmatrix}
5 & 12 \\
21 & 32
\end{bmatrix},
\]

whereas the command without the dot gives the result

\[
\begin{bmatrix}
1 & 2 \\
3 & 4
\end{bmatrix}
\times
\begin{bmatrix}
5 & 6 \\
7 & 8
\end{bmatrix}
= \begin{bmatrix}
19 & 22 \\
43 & 50
\end{bmatrix}.
\]

Similarly, functions can be applied component-wise, e.g.,

\[
\begin{bmatrix}
1 & 2 \\
3 & 4
\end{bmatrix}
\cdot^2
= \begin{bmatrix}
1 & 4 \\
9 & 16
\end{bmatrix}, \quad \cos \left( \begin{bmatrix}
0 & \pi/2 \\
\pi & 2\pi
\end{bmatrix} \right)
= \begin{bmatrix}
1 & 0 \\
-1 & 1
\end{bmatrix}.
\]

Linear systems of equations can be solved with the backslash operator, e.g.,

\[
\begin{bmatrix}
2 & 1 \\
1 & 2
\end{bmatrix}
\backslash
\begin{bmatrix}
3 \\
3
\end{bmatrix}
= \begin{bmatrix}
1 \\
1
\end{bmatrix}
\iff
\begin{bmatrix}
2 & 1 \\
1 & 2
\end{bmatrix}
\ast
\begin{bmatrix}
1 \\
1
\end{bmatrix}
= \begin{bmatrix}
3 \\
3
\end{bmatrix}.
\]
Table B.1  Elementary matrix constructions and operations

<table>
<thead>
<tr>
<th>Expression</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>([a,b,\ldots;x,y,\ldots])</td>
<td>Definition of a matrix (commas may be omitted)</td>
</tr>
<tr>
<td>([a,b,\ldots],[x;y,\ldots])</td>
<td>Definition of row and column vectors</td>
</tr>
<tr>
<td>(A(i,j), I(j))</td>
<td>Entry at position ((i,j)), (j)-th entry</td>
</tr>
<tr>
<td>(a:b, a:step:b)</td>
<td>List of numbers</td>
</tr>
<tr>
<td>(A(i,:), A(:,j))</td>
<td>(i)-th row and (j)-th column</td>
</tr>
<tr>
<td>(A(I,J))</td>
<td>Submatrix defined by lists (I) and (J)</td>
</tr>
<tr>
<td>(\text{ones}(m,n), \text{zeros}(m,n))</td>
<td>Matrix with entries one or zero</td>
</tr>
<tr>
<td>(A+B, A-B, A*B)</td>
<td>Sum, difference, and product</td>
</tr>
<tr>
<td>(A', \text{inv}(A), \text{det}(A))</td>
<td>Transpose, inverse, and determinant</td>
</tr>
<tr>
<td>(x = A\backslash b)</td>
<td>Solution of a linear system of equations</td>
</tr>
<tr>
<td>(\text{eye}(n), \text{speye}(n))</td>
<td>Unit and sparse unit (n \times n) matrix</td>
</tr>
<tr>
<td>(A.*B, A./B)</td>
<td>Component-wise multiplication and division</td>
</tr>
<tr>
<td>(\text{lu}(A), \text{chol}(A))</td>
<td>(LU) and Cholesky factorization</td>
</tr>
<tr>
<td>(\text{eig}(A), \text{diag}(A), \text{tril}(A))</td>
<td>Eigenvectors and eigenvalues, diagonal and lower triangular part</td>
</tr>
<tr>
<td>(\text{sparse}(I,J,X,m,n))</td>
<td>Creation of a sparse matrix</td>
</tr>
</tbody>
</table>

The backslash operator is flexible and can also be used, e.g., to solve overdetermined or singular systems in an appropriate sense. To solve large linear systems of equations with sparse system matrices, it is important that the matrices are defined correspondingly. In most cases this can be done using the MATLAB command `sparse`, which generates a matrix by providing the coordinates and values of the relevant entries, e.g.,

\[
A = \begin{bmatrix}
1 & 0 & 0 & 0 \\
2 & 3 & 0 & 0 \\
0 & 0 & 4 & 0 \\
0 & 0 & 0 & 5
\end{bmatrix} \quad \iff \quad I = [1, 2, 2, 3, 3, 4], \\
J = [1, 1, 2, 3, 3, 4], \\
X = [1, 2, 3, 2, 2, 5], \\
A = \text{sparse}(I,J,X,4,4). \\
\]

Note that in the example the position \((i,j) = (3,3)\) occurs twice with value 2. By convention, values in the lists corresponding to the same position in the matrix are added. This is a crucial feature for the efficient assembly of finite element matrices.

### B.1.2 List Manipulation

The manipulation of lists and arrays is frequently used in the implementation of finite element methods, e.g., to extract implicit information about a triangulation from the lists of vertices and elements. The arrays `cn`, `n4e`, `Db`, and `Nb` specify the triangulation, e.g., of \(\Omega = (0,1)^2\) with \(I_D = [0,1] \times \{0\} \cup \{1\} \times [0,1]\) and
\[ \Gamma_N = \partial \Omega \setminus \Gamma_D \] into two triangles via

\[
\begin{bmatrix}
0 & 0 \\
1 & 0 \\
1 & 1 \\
0 & 1
\end{bmatrix}, \quad \begin{bmatrix}
1 & 2 & 3 \\
1 & 3 & 4
\end{bmatrix}, \quad \begin{bmatrix}
1 & 2 \\
2 & 3
\end{bmatrix}, \quad \begin{bmatrix}
3 & 4
\end{bmatrix}.
\]

We then obtain the nodes belonging to the Dirichlet boundary via the command

\[ \text{dNodes} = \text{unique} (\text{Db}) \].

Precise information about the sides in the triangulation can be obtained via first arranging all sides of elements in one array. Interior sides then occur twice while boundary sides occur only once. To obtain a list in which all sides of a two-dimensional triangulation only appear once, we use the following commands:

```matlab
all_sides = [n4e(:,[1,2]);n4e(:,[2,3]);n4e(:,[3,1])];
[sides,i,j] = unique(sort(all_sides,2),'rows');
```

In the above example we have

\[
\begin{bmatrix}
1 & 2 \\
1 & 3 \\
2 & 3 \\
3 & 4 \\
3 & 1 \\
4 & 1
\end{bmatrix}, \quad \begin{bmatrix}
1 & 2 \\
1 & 3 \\
1 & 4 \\
2 & 3 \\
3 & 4
\end{bmatrix}.
\]

The output arguments \( i \) and \( j \) of the command `unique` provide mappings between the two arrays and thereby specify sides occurring repeatedly. Table B.2 displays further MATLAB commands for manipulating lists.

### Table B.2  Elementary list manipulation commands

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>sort(A,p)</code></td>
<td>Sorts columns or rows of a matrix</td>
</tr>
<tr>
<td><code>unique(A, 'rows')</code></td>
<td>Extracts unique rows</td>
</tr>
<tr>
<td><code>reshape(A,p,q)</code></td>
<td>Rearranges entries of ( A ) in a ( p \times q ) matrix</td>
</tr>
<tr>
<td><code>A(:)</code></td>
<td>Writes columns of ( A ) as one column vector</td>
</tr>
<tr>
<td><code>repmat(A,r,s)</code></td>
<td>Builds block matrix with copies of ( A )</td>
</tr>
<tr>
<td><code>accumarray(N,X,[m,n])</code></td>
<td>Creates a matrix by summing entries of ( X )</td>
</tr>
<tr>
<td><code>length(x),size(A)</code></td>
<td>Length and dimensions of arrays</td>
</tr>
<tr>
<td><code>setdiff(A,B)</code></td>
<td>Set-theoretic difference of ( A ) and ( B )</td>
</tr>
<tr>
<td><code>max(x),min(x)</code></td>
<td>Minimal and maximal entry</td>
</tr>
<tr>
<td><code>find(I)</code></td>
<td>Indices of nonvanishing entries in ( I )</td>
</tr>
</tbody>
</table>
The right-hand side $f$ in a partial differential equation gives rise to a vector $b \in \mathbb{R}^M$ with entries corresponding to the nodes $(z_1, z_2, \ldots, z_M)$ defined by

$$\int_{\Omega} f \varphi_{z_i} \, dx \approx \sum_{T \in \mathcal{T}; z_i \in T} \frac{|T|}{d+1} f(x_T) = b_i$$

for $i = 1, 2, \ldots, M$. The $\ell$-th row of the array $n4e$ specifies those nodes $z_i$ that belong to the element $T_\ell$. With the help of the command $\text{accumarray}$ the vector $b$ can thus be assembled with the following lines:

$$Z = \frac{1}{(d+1)} \cdot \text{Vol}_T \cdot f(Mp_T); \quad ZZ = \text{repmat}(Z, 1, d+1);$$

$$b = \text{accumarray}(n4e(:,), ZZ(:,), [nC, 1]);$$

Here, $\text{Vol}_T$ and $\text{Mp}_T$ are arrays that contain the volumes and midpoints of the elements.

### B.1.3 Graphics

Finite element functions can be visualized as graphs or color plots with the commands $\text{trisurf}$, $\text{trimesh}$, and $\text{tetramesh}$, e.g., if $u$ is the coefficient vector of a $P1$-finite element function via:

$$\text{trisurf}(n4e, c4n(:,1), c4n(:,2), u);$$

$$\text{tetramesh}(n4e, c4n, u);$$

Other useful commands that plot vector fields as arrows or change the view of an object are listed in Table B.3.

<table>
<thead>
<tr>
<th>Table B.3</th>
<th>MATLAB commands that generate and manipulate plots and figures</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>figure</code></td>
<td>Selects a figure window</td>
</tr>
<tr>
<td><code>plot</code>, <code>plot3</code></td>
<td>Plots a polygonal curve in $\mathbb{R}^2$ or $\mathbb{R}^3$</td>
</tr>
<tr>
<td><code>loglog</code></td>
<td>Plot with logarithmic scaling for both axes</td>
</tr>
<tr>
<td><code>legend</code></td>
<td>Adds a legend to a plot</td>
</tr>
<tr>
<td><code>hold on/off</code></td>
<td>Plotting of several objects in one figure</td>
</tr>
<tr>
<td><code>trisurf, trimesh, tetramesh</code></td>
<td>Displays a triangulation in $\mathbb{R}^2$ or $\mathbb{R}^3$</td>
</tr>
<tr>
<td><code>quiver, quiver3</code></td>
<td>Plots a two- or three-dimensional vector field</td>
</tr>
<tr>
<td><code>drawnow, clf</code></td>
<td>Updates and clears a figure</td>
</tr>
<tr>
<td><code>axis</code></td>
<td>Sets the axes in a figure including color range</td>
</tr>
<tr>
<td><code>axis on/off</code></td>
<td>Switches coordinate axes on or off</td>
</tr>
<tr>
<td><code>xlabel, ylabel</code></td>
<td>Adds labels to axes</td>
</tr>
<tr>
<td><code>colorbar</code></td>
<td>Displays a color bar</td>
</tr>
<tr>
<td><code>subplot</code></td>
<td>Shows several plots in one figure</td>
</tr>
<tr>
<td><code>view</code></td>
<td>Changes the perspective</td>
</tr>
<tr>
<td><code>colormap</code></td>
<td>Chooses a color scale</td>
</tr>
<tr>
<td><code>clc</code></td>
<td>Clears the command window</td>
</tr>
</tbody>
</table>
Table B.4 Standard programming commands

<table>
<thead>
<tr>
<th>Command</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>for . . end</td>
<td>For loop</td>
</tr>
<tr>
<td>while . . end</td>
<td>While loop</td>
</tr>
<tr>
<td>if . . else . . end</td>
<td>If-then-else structure</td>
</tr>
<tr>
<td>pause, break</td>
<td>Pause until key is hit, stop a program</td>
</tr>
<tr>
<td>disp, fprintf</td>
<td>Displays or prints a quantity</td>
</tr>
<tr>
<td>load, save</td>
<td>Loads and saves variables</td>
</tr>
<tr>
<td>fopen, fclose</td>
<td>Opens and closes file to write data</td>
</tr>
<tr>
<td>addpath, rmpath</td>
<td>Adds and removes a directory path</td>
</tr>
<tr>
<td>Ctrl-C</td>
<td>Stops a running program</td>
</tr>
</tbody>
</table>

Table B.5 Examples of functions available in MATLAB

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>exp, ln, log</td>
<td>Exponential and logarithms</td>
</tr>
<tr>
<td>sqrt, ^</td>
<td>Square root and power</td>
</tr>
<tr>
<td>sin, cos, tan, pi</td>
<td>Trigonometric functions and constant $\pi$</td>
</tr>
</tbody>
</table>

B.1.4 Standard Commands

Most standard programming structures and mathematical functions are available in MATLAB. Tables B.4 and B.5 provide an overview over frequently used commands.

B.2 Finite Element Matrix Assembly

B.2.1 Global Loops

The standard assembly of a finite element matrix is based on a global loop over elements in a triangulation, e.g., via

$$ s_{zy} = \int_{\Omega} \nabla \varphi_z \cdot \nabla \varphi_y \, dx = \sum_{T \in \mathcal{T}_h} |T| \nabla \varphi_z |_T \cdot \nabla \varphi_y |_T. $$

A realization in MATLAB is shown in Fig. B.1. The routine computes the coordinate lists \( I, J \) and corresponding entries \( X \) of the \( P_1 \)-finite element stiffness matrix, which are then used to assemble the stiffness matrix via

$$ s = \text{sparse}(I,J,X,nC,nC); $$

For the use of the routine, the volumes of elements and elementwise gradients of nodal basis functions are precomputed and stored in the arrays \( \text{Vol}_T \) and \( \text{Grads}_T \) arranged by elements \( T_1, T_2, \ldots, T_L \) and local node numbers \( z_i^T, i = \ldots \).
506 B Implementation Aspects

```matlab
function [I,J,X] = fe_matrix_loop(c4n,n4e,Vol_T,Grads_T)
d = size(c4n,2); nE = size(n4e,1);
ctr = 0; ctr_max = (d+1)^2*nE;
I = zeros(ctr_max,1); J = zeros(ctr_max,1);
X = zeros(ctr_max,1);
for j = 1:nE
    grads_T = Grads_T((j-1)*(d+1)+(1:d+1),:);
    vol_T = Vol_T(j);
    for m = 1:d+1
        for n = 1:d+1
            ctr = ctr+1; I(ctr) = n4e(j,m); J(ctr) = n4e(j,n);
            X(ctr) = vol_T*grads_T(m,:)*grads_T(n,:)';
        end
    end
end
end
```

**Fig. B.1** Standard assembly of the stiffness matrix in a loop over all elements

\[
1, 2, \ldots, d+1, \text{i.e.,}
\]

\[
\begin{bmatrix}
|T_1| \\
|T_2| \\
\vdots \\
|T_L|
\end{bmatrix} \in \mathbb{R}^L, \quad \text{Grads}_T = \begin{bmatrix}
\nabla \varphi_{1|T_1} \\
\vdots \\
\nabla \varphi_{L+1|T_1} \\
\nabla \varphi_{1|T_2} \\
\vdots \\
\nabla \varphi_{L+1|T_2} \\
\vdots \\
\nabla \varphi_{1|T_L} \\
\vdots \\
\nabla \varphi_{L+1|T_L}
\end{bmatrix} \in \mathbb{R}^{L(d+1) \times d}.
\]

Although the assembly via a loop over elements is of linear complexity, its practical performance in interpreted programming languages is suboptimal, i.e., its runtime is typically longer than the CPU-time needed for solving the linear system of equations. This is problematic in the case of time-dependent problems, when a system matrix has to be assembled in every time step.

### B.2.2 Vectorized Loop

A way to accelerate the computation of the entries of the stiffness matrix is to avoid the global loop over elements and to compute the relevant quantities with matrix operations. The idea is related to the simple observation that e.g.,

\[
\sum_{i=1}^{n} a(i) \ast b(i) = a' \ast b = \text{sum}(a \ast b),
\]
B.2 Finite Element Matrix Assembly

where the implementation of the method on the right-hand side is significantly faster than the loop suggested by the left-hand side in a MATLAB implementation. The first code displayed in Fig. B.2 results from the routine shown in Fig. B.1 by eliminating the global loop and writing corresponding operations in vectorized form. In the second code of Fig. B.2, the loops with boundedly many repetitions have also been eliminated. These implementations are significantly faster than the code that uses the loop over elements, which is due to the optimized implementation of matrix operations in MATLAB.

B.2.3 Assembly in C

An alternative to the acceleration via vectorization is the assembly of the matrix entries in the compiled programming language C. To include C code in MATLAB, we use the interface MEX, which allows for a simple variable transfer. Figure B.3 shows a C code that is equivalent to the MATLAB routine shown in Fig. B.1. Its compilation is done in MATLAB using the command

```
mex fe_matrix_mex.c;
```
#include <mex.h>  /* fe_matrix_mex.c */
void lists(double n4e[], double c4n[],
    double Vol_T[], double Grads_T[],
    int nE, int nC, int d,
    double I[], double J[], double X[]){
    int j, m, n, r, idx1, idx2, ctr;
    double val;
    ctr = 0;
    for (j=0; j<nE; j++){
        for (m=0; m<d+1; m++){
            for (n=0; n<d+1; n++){
                I[ctr] = n4e[j+m*nE]; J[ctr] = n4e[j+n*nE];
                val = 0.0;
                for (r=0; r<d; r++){
                    idx1 = j*(d+1)+m+r*(d+1)*nE;
                    idx2 = j*(d+1)+n+r*(d+1)*nE;
                    X[ctr] += Vol_T[j]*Grads_T[idx1]*Grads_T[idx2];
                }
                ctr += 1;
            }
        }
    }
}
void mexFunction(int nlhs, mxArray *plhs[], int nrhs,
    const mxArray *prhs[]){
    double *n4e, *c4n, *Vol_T, *Grads_T;
    int nE, nC, d;
    double *I, *J, *X;
    if (nrhs != 4)
        mexErrMsgTxt("4 input arguments required");
    nC = mxGetM(prhs[0]);
    d = mxGetN(prhs[0]);
    nE = mxGetM(prhs[1]);
    c4n = mxGetPr(prhs[0]);
    n4e = mxGetPr(prhs[1]);
    Vol_T = mxGetPr(prhs[2]);
    Grads_T = mxGetPr(prhs[3]);
    if (nlhs != 3)
        mexErrMsgTxt("3 output arguments required");
    plhs[0] = mxCreateDoubleMatrix(nE*(d+1)*(d+1),1,mxREAL);
    plhs[1] = mxCreateDoubleMatrix(nE*(d+1)*(d+1),1,mxREAL);
    plhs[2] = mxCreateDoubleMatrix(nE*(d+1)*(d+1),1,mxREAL);
    I = mxGetPr(plhs[0]);
    J = mxGetPr(plhs[1]);
    X = mxGetPr(plhs[2]);
    lists(n4e,c4n,Vol_T,Grads_T,nE,nC,d,I,J,X);
}

Fig. B.3  Computing the entries of the stiffness matrix in C using MATLAB to C interface MEX

For this the gnu C compiler gcc has to be selected via the MATLAB command mex -setup. The routine can then be called within MATLAB as in the code shown in Fig. B.4.
function p1_comparison(d,red,assembly,solver)
    [c4n,n4e,Db,Nb] = triang_cube(d); Db = [Db;Nb]; Nb = [];
    for j = 1:red
        [c4n,n4e,Db,Nb] = red_refine(c4n,n4e,Db,Nb);
    end
    nC = size(c4n,1); h = 1/nC^(1/d);
    dNodes = unique(Db); fNodes = setdiff(1:nC,dNodes);
    [Vol_T,Grads_T,Mp_T] = nodal_basis(c4n,n4e);
    Z = (1/((d+1)))*Vol_T.*f(Mp_T); ZZ = repmat(Z,1,d+1);
    switch assembly
        case 0
            [I,J,X] = fe_matrix_loop(c4n,n4e,Vol_T,Grads_T);
        case 1
            [I,J,X] = fe_matrix_vectorized_1(c4n,n4e,Vol_T,Grads_T);
        case 2
            [I,J,X] = fe_matrix_vectorized_2(c4n,n4e,Vol_T,Grads_T);
        case 3
            [I,J,X] = fe_matrix_mex(c4n,n4e,Vol_T,Grads_T);
    end
    s = sparse(I,J,X,nC,nC); u = u_D(c4n);
    b = accumarray(n4e(:),ZZ(:),[nC,1])-s*u;
    s_fN = s(fNodes,fNodes); b_fN = b(fNodes);
    switch solver
        case 0
            u(fNodes) = s_fN\b_fN;
        case 1
            K = size(fNodes,2); eps_stop = h;
            C = spdiags(diag(s_fN),0,K,K);
            u(fNodes) = pcg(s_fN,b_fN,eps_stop,K,C);
    end
    show_p1(c4n,n4e,Db,Nb,u);
    function val = f(x); val = ones(size(x,1),1);
    function val = u_D(x); val = zeros(size(x,1),1);
    function [Vol_T,Grads_T,Mp_T] = nodal_basis(c4n,n4e)
        d = size(c4n,2); nE = size(n4e,1);
        Grads_T = zeros((d+1)*nE,d);
        Vol_T = zeros(nE,1); Mp_T = zeros(nE,d);
        for j = 1:nE
            X_T = [ones(1,d+1);c4n(n4e(j,:),:);]';
            Grads_T((j-1)*(d+1)+(1:d+1),:) = X_T\[zeros(1,d);eye(d)];
            Vol_T(j) = det(X_T)/factorial(d);
            Mp_T(j,:) = sum(c4n(n4e(j,:),:),1)/(d+1);
        end

Fig. B.4  Comparison of four different ways of computing the entries of the P1-finite element stiffness matrix and two different solvers for the linear system of equations.
B.2.4 Comparison

Tables B.6 and B.7 show a comparison of the runtimes for the different ways to assemble the stiffness matrix for two- and three-dimensional Poisson problems:

$$-\Delta u = 1 \text{ in } \Omega = (0, 1)^d, \quad u = 0 \text{ on } \Gamma_D = \partial \Omega.$$  

The experiments were carried out on a standard Desktop (Intel Core i3-3220 CPU, 8 GB RAM). To relate the numbers, we also included the runtimes of different solvers for the linear systems of equations, where $A \backslash b$ refers to the solution via backslash operator and PCG to the use of a preconditioned conjugate gradient algorithm with the diagonal Jacobi preconditioner. A dash indicates that an experiment was not carried out due to memory limitations.

Our conclusions from the experiments are as follows:

- all assembly routines scale linearly with respect to the numbers of degrees of freedom, i.e., when the mesh-size is halved, the runtime increases approximately by a factor $2^d$;
- a vectorized assembly reduces the runtime to a few percent of the assembly via a loop in MATLAB and is comparable to the solution with the backslash operator;
- eliminating loops with a mesh-independent number of repetitions does not lead to an additional runtime reduction;

Table B.6 Runtime comparison for different assemblies of the stiffness matrix and solution methods for a two-dimensional Poisson problem; numbers are in seconds

<table>
<thead>
<tr>
<th>$d = 2$, $\mathcal{N}_h$</th>
<th>M-loop</th>
<th>vec-1</th>
<th>vec-2</th>
<th>C-loop</th>
<th>$A \backslash b$</th>
<th>PCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>0.0007</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0020</td>
</tr>
<tr>
<td>25</td>
<td>0.0021</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0024</td>
</tr>
<tr>
<td>81</td>
<td>0.0076</td>
<td>0.0003</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0029</td>
</tr>
<tr>
<td>289</td>
<td>0.0148</td>
<td>0.0002</td>
<td>0.0002</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0025</td>
</tr>
<tr>
<td>1089</td>
<td>0.0590</td>
<td>0.0007</td>
<td>0.0007</td>
<td>0.0002</td>
<td>0.0011</td>
<td>0.0053</td>
</tr>
<tr>
<td>4225</td>
<td>0.2411</td>
<td>0.0026</td>
<td>0.0023</td>
<td>0.0006</td>
<td>0.0054</td>
<td>0.0218</td>
</tr>
<tr>
<td>16641</td>
<td>0.9608</td>
<td>0.0137</td>
<td>0.0099</td>
<td>0.0028</td>
<td>0.0284</td>
<td>0.1123</td>
</tr>
<tr>
<td>66049</td>
<td>3.9347</td>
<td>0.0651</td>
<td>0.0397</td>
<td>0.0107</td>
<td>0.1551</td>
<td>0.8683</td>
</tr>
<tr>
<td>263169</td>
<td>15.8410</td>
<td>0.3031</td>
<td>0.2444</td>
<td>0.0571</td>
<td>0.7964</td>
<td>7.8872</td>
</tr>
</tbody>
</table>

Table B.7 Runtime comparison for different assemblies of the stiffness matrix and solution methods for a three-dimensional Poisson problem; numbers are in seconds

<table>
<thead>
<tr>
<th>$d = 3$, $\mathcal{N}_h$</th>
<th>M-loop</th>
<th>vec-1</th>
<th>vec-2</th>
<th>C-loop</th>
<th>$A \backslash b$</th>
<th>PCG</th>
</tr>
</thead>
<tbody>
<tr>
<td>27</td>
<td>0.0045</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0129</td>
<td>0.0000</td>
<td>0.3710</td>
</tr>
<tr>
<td>125</td>
<td>0.0193</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0001</td>
<td>0.0003</td>
<td>0.0134</td>
</tr>
<tr>
<td>729</td>
<td>0.1516</td>
<td>0.0026</td>
<td>0.0027</td>
<td>0.0006</td>
<td>0.0020</td>
<td>0.0019</td>
</tr>
<tr>
<td>4913</td>
<td>1.2204</td>
<td>0.0273</td>
<td>0.0218</td>
<td>0.0044</td>
<td>0.0439</td>
<td>0.0077</td>
</tr>
<tr>
<td>35937</td>
<td>10.0983</td>
<td>0.2363</td>
<td>0.1851</td>
<td>0.0345</td>
<td>2.3275</td>
<td>0.0979</td>
</tr>
<tr>
<td>274625</td>
<td>–</td>
<td>2.0396</td>
<td>1.6933</td>
<td>0.3291</td>
<td>–</td>
<td>1.7406</td>
</tr>
</tbody>
</table>
• a loop in C reduces the assembly runtime to about ten percent of the solution
time;
• the backslash operator provides good results in two-dimensional situations but
may fail in three-dimensional situations;
• in the three-dimensional situation, the preconditioned conjugate gradient algo-
rithm with Jacobi preconditioner outperforms the solution via the backslash
operator;
• the solution of the linear system of equations is the bottleneck of the problem,
i.e., dominates the total CPU time, provided that the assembly of the stiffness
matrix is done appropriately.

Remark B.1 Precomputation of the elementwise gradients of the nodal basis func-
tions in the routine nodal_basis.m in Fig. B.4 can be accelerated by vectorizing
the computation or transferring it to C. A useful formula for efficient computation
of the gradients is the identity
\[
\nabla \varphi_i|_T = -\frac{|S_z|}{d|T|} n_{S_z},
\]
for \(d = 2, 3\), where \(S_z\) is the side of \(T\) which is opposite to \(z\) and \(n_{S_z}\) is the outward
unit normal on \(S_z\). The quantity \((d - 1)|S_z|n_{S_z}\) can be obtained from a cross-product
of edge vectors if \(d = 3\), or a rotation by \(\pi/2\) of an edge vector if \(d = 2\).

B.3 Mesh Generation and Visualization

B.3.1 Mesh Generation

The free MATLAB package distmesh provides routines to generate triangulations
of domains in \(\mathbb{R}^d\). It assumes the domain to be defined by a signed distance function
\(s_\Omega : \mathbb{R}^d \rightarrow \mathbb{R}\) and contained in a box of diameter \(2R\), i.e,
\[
\Omega = \{x \in \mathbb{R}^d : s_\Omega(x) < 0\} \subset K_R^\infty(0),
\]
where \(K_R = \{x \in \mathbb{R}^d : \|x\|_\infty \leq R\}\). It is assumed that \(s_\Omega\) grows linearly away from
\(\partial \Omega\), i.e., that
\[
|\nabla s_\Omega| = 1.
\]
On \(\partial \Omega\) the gradient \(\nabla s_\Omega\) coincides with the outer unit normal to \(\Omega\). A triangulation
is generated with the command:
\[
[c4n,n4e] = \text{distmeshnd}(\text{@sdist}_\text{Omega}, \text{@mesh_density}, \ldots \\
\text{h_min, bd_box, fixed_vertices});
\]
Fig. B.5 Generation of a graded triangulation of a $d$-dimensional ball $B_r(0)$

The parameter $h_{\text{min}}$ determines the approximate minimal mesh-size $h_{\text{min}}$ of the triangulation. The function $\text{mesh}\_\text{density}$ allows for a grading of the triangulation specified by a function $\phi(x)$ so that

$$\text{diam}(T) \approx \phi(x_T) h_{\text{min}}$$

for all triangles or tetrahedra $T$ with midpoint $x_T$ in the generated triangulation $\mathcal{T}_h$. The argument $\text{fixed}\_\text{vertices}$ allows for prescribing vertices in the triangulation.

Example B.1 The generation of a triangulation of the ball $B_r(x) \subset \mathbb{R}^d$ with radius $r > 0$ using the function $s_\Omega(x) = |x| - r$, the bounding box defined by $R = r$, minimal mesh-size $h_{\text{min}}$, and grading function $\phi(x) = (1+|x|)^{\alpha}$ is shown in Fig. B.5.

The function $\text{distmeshnd}$ initially chooses a quasiuniform triangulation of the bounding box $K_\varepsilon(0)$, discards all vertices that do not belong to $\Omega$, i.e., vertices $z$ for which $s_\Omega(z) > \varepsilon_{\text{tol}}$, and then equidistributes and projects the remaining vertices with the help of a repulsive force function that involves the function $\phi$. A large class of Lipschitz domains $\Omega$ can be obtained as the union or intersection of simple domains and their complements. If $\Omega_1$ and $\Omega_2$ are Lipschitz domains with signed distance functions $s_{\Omega_1}$ and $s_{\Omega_2}$, then signed distance functions for the union, intersection, complement, and transformed domain are given by

$$s_{\Omega_1 \cup \Omega_2} = \min\{s_{\Omega_1}, s_{\Omega_2}\},$$

$$s_{\Omega_1 \cap \Omega_2} = \max\{s_{\Omega_1}, s_{\Omega_2}\},$$

$$s_{\mathbb{R}^d \setminus \Omega_1} = -s_{\Omega_1},$$

$$s_{\phi(\Omega_1)} = s_{\Omega_1} \circ \phi^{-1}.$$
Fig. B.6 Construction of a signed distance function for a parallelepiped

A signed distance function for the parallelepiped $\Omega = \prod_{i=1}^d (-\ell_i, \ell_i)$ is obtained from the functions

$$s_i(x) = |x_i| - \ell_i$$

via $s_\Omega = \max_{i=1,2,\ldots,d} s_i$. The performance of the routine `distmeshnd` is significantly improved by setting

$$s_\Omega(x) = \begin{cases} \left( \sum_{i=1}^d s_i(x)^2 \right)^{1/2}, & s_i(x) > 0 \text{ for } i = 1, 2, \ldots, d, \\ \max_{i=1,2,\ldots,d} s_i(x), & \text{otherwise}, \end{cases}$$

cf. Fig. B.6. With this definition the function $s_\Omega(x)$ is a proper distance function to $\Omega$ in the sense that $s_\Omega(x) = \inf_{y \in \Omega} |x - y|$ for all $x \in \mathbb{R}^d \setminus \Omega$. This function provides a mechanism to compute projections onto $\partial \Omega$. Alternatively, boundary points of $\Omega$ can be prescribed as fixed vertices to improve the stability of the mesh generation routine.

**Example B.2** Figure B.7 shows a function that generates a triangulation of a cylinder domain of length $2L_c$ and diameter $2r_c$, with a spherical hole of diameter $2r_s$ centered at the origin, i.e.,

$$\Omega = \left( (-L_c, L_c) \times B_{r_c}(0) \right) \setminus B_{r_s}(0) \subset \mathbb{R}^d.$$  

The subroutines `fix_orientation` and `find_bdy_sides` adjust the ordering of the nodes of triangles or tetrahedra to obtain a positive orientation and identify the boundary sides in the triangulation.
function gen_triang_cyl_w_hole(d_tmp)
addpath '~/auxiliary/distmesh';
global d r_sph L_cyl r_cyl;
d = d_tmp; L_cyl = 2; r_cyl = 1; r_sph = 1/2;
R = 2; h_min = 0.1; fixed = [];
box = [-R*ones(1,d);R*ones(1,d)];
[c4n,n4e] = distmeshnd(@s,@phi,h_min,box,fixed);
str = strcat('save triang_cyl_w_hole_','d.mat c4n n4e Db Nb');
end

function val = s(x)
global d r_sph L_cyl r_cyl;
dist_hor = abs(x(:,1))-L_cyl;
dist_rad = sqrt(sum(x(:,2:d).^2,2))-r_cyl;
dist_cyl = max(dist_hor,dist_rad);
dist_compl_sph = r_sph-sqrt(sum(x.^2,2));
val = max(dist_cyl,dist_compl_sph);

function val = phi(x)
global r_sph;
dist_sph = sqrt(sum(x.^2,2))-r_sph;
val = min(dist_sph+1,2);

function bdy = find_bdy_sides(n4e)
d = size(n4e,2)-1;
if d == 2
    all_sides = [n4e(:,[1,2]);n4e(:,[2,3]);n4e(:,[3,1])];
elseif d == 3
    all_sides = [n4e(:,[2,4,3]);n4e(:,[1,3,4]);n4e(:,[1,4,2]);n4e(:,[1,2,3])];
end
[sides,-,j] = unique(sort(all_sides,2),'rows');
valence = accumarray(j(:),1);
bdy = sides(valence==1,:);

function n4e = fix_orientation(c4n,n4e)
global d; nE = size(n4e,1); or_Vol_T = zeros(nE,1);
for j = 1:nE
    X_T = [ones(1,d+1);c4n(n4e(j,:),:)'];
    or_Vol_T(j) = det(X_T)/factorial(d);
end
n4e(or_Vol_T<0,[1,2]) = n4e(or_Vol_T<0,[2,1]);

Fig. B.7 Generation of a triangulation of a cylindrical domain with hole
B.3.2 Visualization

To generate plots with streamlines of velocity fields or to produce movies from a sequence of plots of finite element functions, the platform-independent and free application paraview, cf. Fig. B.8, can be used. It requires the vtu file format of the triangulation and finite element functions. Finite element functions can be elementwise constant functions, specified by cell values, or continuous elementwise affine functions, specified by point values. Functions can be scalar-valued or vector fields. The data file shown in Fig. B.9 was generated with the MATLAB routine export2vtu.m displayed in Fig. B.10 for functions and vector fields specified in the following example. All vectorial quantities are embedded into $\mathbb{R}^3$ by appending a trivial component if necessary.

**Example B.3** Let $\Omega = (0, 1)^2$ and $\mathcal{T}_h$ be the triangulation of $\Omega$ consisting of two triangles specified by

$$
\begin{align*}
z_1 &= (0, 0), \\
z_2 &= (1, 0), \\
z_3 &= (1, 1), \\
z_4 &= (0, 1)
\end{align*}
$$

and

$$
\begin{align*}
T_1 &= \text{conv}\{z_1, z_2, z_3\}, \\
T_2 &= \text{conv}\{z_1, z_3, z_4\}.
\end{align*}
$$
Fig. B.9 A vtu file that specifies a triangulation and finite element functions (displayed using Opera)

Let the scalar and vectorial affine functions $p_h^1 \in \mathcal{S}^1(\mathcal{T}_h)$ and $u_h^1 \in \mathcal{S}^1(\mathcal{T}_h)^2$ be defined by

\[
p_h^1(z_i) = i, \quad u_h^1(z_i) = 2(i - 1) + \begin{bmatrix} 5 \\ 6 \end{bmatrix}
\]

for $i = 1, 2, \ldots, 4$. Let the scalar and vectorial elementwise constant functions $p_h^0 \in \mathcal{L}^0(\mathcal{T}_h)$ and $u_h^0 \in \mathcal{L}^0(\mathcal{T}_h)^2$ be defined by

\[
p_h^0|_{T_i} = 100 + i, \quad u_h^0|_{T_i} = 100 + 2(i - 1) + \begin{bmatrix} 3 \\ 4 \end{bmatrix}
\]
function export2vtu(file,c4n,n4e,p_p1,u_p1,p_p0,u_p0)
[nC,d] = size(c4n); nE = size(n4e,1); type = d^2+1;
if d == 2
    c4n = [c4n, zeros(nC,1)];
    u_p1 = [u_p1(1:2:2*nC), u_p1(2:2:2*nC), zeros(nC,1)];
    u_p0 = [u_p0(1:2:2*nE), u_p0(2:2:2*nE), zeros(nE,1)];
end
fid = fopen(fullfile(file,'.vtu'),'wt');
fprintf(fid,’<?xml version=“1.0”?>
’);
fprintf(fid,’<VTKFile type=“UnstructuredGrid” version=“0.1”>
’);
fprintf(fid,’<UnstructuredGrid>
’);
fprintf(fid,’<Piece NumberOfPoints=“%d” NumberOfCells=“%d”>
’,nC,nE);
fprintf(fid,’<PointData>
’);
fprintf(fid,’<DataArray type=“Float32” Name=“p_p1”>
’);
fprintf(fid,’%3.3f
’,p_p1);
fprintf(fid,’</DataArray>
’);
fprintf(fid,’<DataArray type=“Float32” Name=“u_p1” NumberOfComponents=“3”>
’);
fprintf(fid,’%3.3f
’,u_p1);
fprintf(fid,’</DataArray>
’);
fprintf(fid,’</PointData>
’);
fprintf(fid,’<CellData>
’);
fprintf(fid,’<DataArray type=“Float32” Name=“p_p0”>
’);
fprintf(fid,’%3.3f
’,p_p0);
fprintf(fid,’</DataArray>
’);
fprintf(fid,’<DataArray type=“Float32” Name=“u_p0” NumberOfComponents=“3”>
’);
fprintf(fid,’%3.3f
’,u_p0);
fprintf(fid,’</DataArray>
’);
fprintf(fid,’</CellData>
’);
fprintf(fid,’<Points>
’);
fprintf(fid,’<DataArray type=“Float32” NumberOfComponents=“3”>
’);
fprintf(fid,’%3.3f
’,c4n);
fprintf(fid,’</DataArray>
’);
fprintf(fid,’</Points>
’);
fprintf(fid,’<Cells>
’);
fprintf(fid,’<DataArray type=“Int32” Name=“connectivity”>
’);
fprintf(fid,’%d
’,n4e-1);
fprintf(fid,’</DataArray>
’);
fprintf(fid,’<DataArray type=“Int32” Name=“offsets”>
’);
fprintf(fid,’%d
’,d+1:d+1:(d+1)*nE);
fprintf(fid,’</DataArray>
’);
fprintf(fid,’<DataArray type=“UInt8” Name=“types”>
’);
fprintf(fid,’%d
’,type*ones(nE,1));
fprintf(fid,’</DataArray>
’);
fprintf(fid,’</Cells>
’);
fprintf(fid,’</UnstructuredGrid>
’);
fprintf(fid,’</VTKFile>
’);
fclose(fid);

Fig. B.10 Export of a triangulation and scalar and vectorial finite element functions to vtu format
Table B.8  Filters and source functions for visualizing finite element functions in *paraview*

<table>
<thead>
<tr>
<th>Filter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>slice</td>
<td>Displays cut through domain</td>
</tr>
<tr>
<td>clip</td>
<td>Displays cross section of domain</td>
</tr>
<tr>
<td>glyph</td>
<td>Visualizes velocity field</td>
</tr>
<tr>
<td>warp</td>
<td>Visualizes displacement field</td>
</tr>
<tr>
<td>contour</td>
<td>Visualizes level sets</td>
</tr>
<tr>
<td>stream</td>
<td>Displays stream lines</td>
</tr>
<tr>
<td>annotate time</td>
<td>Displays a time counter</td>
</tr>
<tr>
<td>calculator</td>
<td>Allows for modifying, e.g., scaling, variables</td>
</tr>
<tr>
<td>text (source)</td>
<td>Displays text</td>
</tr>
</tbody>
</table>

for \(i = 1, 2\). A *vtu* file that specifies the triangulation and the functions is shown in Fig. B.9. The file was generated with the MATLAB routine `export2vtu.m` shown in Fig. B.10 via the following commands:

```matlab
>> c4n = [0 0;1 0;1 1;0 1]; n4e = [1 2 3;1 3 4];
>> p_p1 = [1;2;3;4]; u_p1 = [5;6;7;8;9;10;11;12];
>> p_p0 = [101;102]; u_p0 = [103;104;105;106];
>> file = 'test';
>> export2vtu(file,c4n,n4e,p_p1,u_p1,p_p0,u_p0);
```

### B.3.3 Manipulating Plots

The visualization of finite element functions or vector fields in *paraview* is done using filters. Some examples are listed in Table B.8. A sequence of data files can be exported as a movie using the file menu item *Save Animation*; background and text colors can be modified in the menu item *Settings*.

### References

Further details on implementing finite element methods in MATLAB can be found in [1–3]. Details about *paraview* are provided in [5], the usage of the mesh generator *distmesh* is explained in [6]. Another powerful mesh generator is *gmsh* which is described in [4].

Appendix C
Notation, Inequalities, Guidelines

C.1 Frequently Used Notation

Real Numbers, Vectors, and Matrices

\[ Z, \mathbb{N}, \mathbb{N}_0 \]
Integers, positive and nonnegative integers

\[ \mathbb{R}, \mathbb{C}, \mathbb{R}_{\geq 0} \]
Real and complex numbers, nonnegative real numbers

\( [s, t], (s, t) \)
Closed and open interval

\( \mathbb{R}^n \)
\( n \)-dimensional Euclidean vector space

\( \mathbb{R}^{m \times n} \)
Vector space of \( m \) by \( n \) matrices

\( B_r(x), B_r \)
Open ball of radius \( r \) centered at \( x \) or at the origin

\( K_r(x), K_r \)
Closed ball of radius \( r \) centered at \( x \) or at the origin

\( A \subseteq B \)
\( A \) is a subset of \( B \) or \( A = B \)

\( a, A \)
(Column) vector and matrix

\( a^\top, A^\top \)
Transpose of a vector or matrix

\( \| \cdot \| \)
Euclidean length or Frobenius norm

\( a \cdot b = a^\top b \)
Scalar product of vectors \( a \) and \( b \)

\( A : B \)
Inner product of matrices \( A \) and \( B \)

\( \text{tr} A \)
Trace of the matrix \( A \)

\( I_L \)
\( L \times L \) identity matrix

\( [x, y]^\top \)
Vectors with entries \( x \) and \( y \)

\[ \begin{bmatrix} x_1 & x_2 \\ y_1 & y_2 \end{bmatrix} \]
Matrix with entries \( x_1, x_2, y_1, y_2 \)

\( a \gg b \)
a significantly greater than \( b \)

\( a \approx b \)
a approximately equal to \( b \)

\( a \sim b \)
a proportional to \( b \)
Sets and Domains

\( \Omega \)  \quad \text{Bounded Lipschitz domain in } \mathbb{R}^d, \ d = 2, 3 \\
\partial \Omega \quad \text{Boundary of the domain } \Omega \\
n \quad \text{Outer unit normal on } \partial \Omega \\
\Gamma_D \quad \text{Dirichlet boundary, closed subset of } \partial \Omega \\
\Gamma_N \quad \text{Neumann boundary, } \Gamma_N = \partial \Omega \setminus \Gamma_D \\
[0, T] \quad \text{Time interval}

Linear Spaces and Operators

id \quad \text{Identity operator} \\
kern \quad \text{Kernel of an operator} \\
X, \ Y \quad \text{Banach spaces} \\
\| \cdot \|_X \quad \text{Norm in } X \\
X' \quad \text{Linear bounded functionals } \phi : X \to \mathbb{R} \\
\langle \phi, x \rangle = \phi(x) \quad \text{Duality pairing of } \phi \in X' \text{ and } x \in X \\
\| \cdot \|_{X'} \quad \text{Operator norm in } X' \\
L(X, Y) \quad \text{Bounded linear operators } A : X \to Y \\
\| \cdot \|_{L(X, Y)} \quad \text{Operator norm in } L(X, Y) \\
A' \quad \text{Adjoint of } A \in L(X, Y) \\
H \quad \text{Hilbert space} \\
(x, y)_H \quad \text{Inner product of } x \text{ and } y \text{ in a Hilbert space } H

Differential Operators

\partial_i, \partial_{x_i}, \frac{\partial}{\partial x_i} \quad \text{Partial derivative with respect to the } i\text{-th coordinate} \\
\nabla \quad \text{Gradient of a function} \\
div \quad \text{Divergence of a vector field} \\
D, D^2 \quad \text{Total derivative and Hessian of a function} \\
\partial_x, \partial_y, \partial_t, \partial^a \quad \text{Partial derivatives} \\
\partial_n u = \nabla u \cdot n \quad \text{Normal derivative on } \partial \Omega \\
u_t \quad \text{Partial derivative with respect to } t \\
\varepsilon(u) \quad \text{Symmetric gradient of a vector field} \\
\Delta \quad \text{Laplace operator}
C.1 Frequently Used Notation

Function Spaces

\[ C^k(A) \] \( k \)-times continuously differentiable functions

\[ C_c^\infty(\Omega) \] Compactly supported, smooth functions

\[ C(\Omega) \] Closure of \( C_c^\infty(\Omega) \) with respect to maximum norm

\[ L^p(\Omega) \] Functions whose \( p \)-th power is Lebesgue integrable

\[ W^{k,p}(\Omega) \] \( k \)-times weakly differentiable functions

\[ W_0^{k,p}(\Omega) \] \( k \)-times weakly differentiable functions

\[ H^k(\Omega) \] Hilbert space \( W^{k,2}(\Omega) \)

\[ H_N(\text{div}; \Omega) \] Vector fields with square integrable divergence

\[ \| \cdot \|, (\cdot, \cdot) \] Norm and inner product in \( L^2(\Omega; \mathbb{R}^m) \)

\[ L^2(\Omega) \] Functions in \( L^2(\Omega) \) with vanishing integral over \( \Omega \)

Finite Differences

\( \tau, \Delta t, \Delta x, \Delta y \) Step-sizes

\( \partial^-, \partial^+, \partial \) Backward, forward, and central difference quotient

\( d_t \) Backward difference quotient in time

\( t_k, t_{k+1/2} \) Time steps \( k\tau \) and \( (k + 1/2)\tau \)

\( u^k, u^{k+1/2} \) Approximations associated with time steps

Finite Element Spaces

\( h, h_{\min} \) Maximal and minimal diameter of elements in \( \mathcal{T}_h \)

\( h_T, h_S, h_z \) Local mesh-sizes

\( N_h \) Nodes that define vertices of elements

\( \mathcal{T}_h \) Set of elements that define a triangulation

\( \mathcal{S}_h \) Sides of elements in a triangulation

\( z, E, S, T \) Node, edge, side, and element in a triangulation

\( \psi_z \) Nodal basis function

\( \omega_z \) Patch of a node

\( P_k(T) \) Polynomials of maximal degree \( k \) restricted to \( T \)

\( L^0(\mathcal{T}_h) \) \( \mathcal{T}_h \)-elementwise constant functions

\( \mathcal{P}_1(\mathcal{T}_h) \) Continuous, \( \mathcal{T}_h \)-elementwise affine functions

\( \mathcal{P}_0(T), \mathcal{P}_0^1(\mathcal{T}_h) \) Functions in \( \mathcal{P}_1(\mathcal{T}_h) \) vanishing on \( \Gamma_D \) or \( \partial \Omega \)

\( \mathcal{I}_h \) Nodal interpolation operator on \( \mathcal{T}_h \)

\( \mathcal{J}_h \) Clément quasi-interpolant

\( [ \nabla u_h \cdot n_S ] \) Jump of the normal component of \( \nabla u_h \) across \( S \)

\( \mathcal{P}^{1,\Sigma}(\mathcal{T}_h) \) Crouzeix–Raviart finite element space
Other Notation

\( c, C, C', C'' \), \( c_1, c_2, \ldots \) Mesh-size independent, generic constants
\( d, \, ds \) Volume and surface element for Lebesgue measure
\( \overline{A} \) Closure of a set \( A \)
\( |A| \) Cardinality, volume, surface area, or length of a set \( A \)
\( \text{diam}(A) \) Diameter of the set \( A \)
\( \chi_A \) Characteristic function of a set \( A \)
\( \delta_{ij} \) Kronecker symbol
\( o(t), O(t) \) Landau symbols
\( \text{supp} f \) Support of a function \( f \)
\( C \) Consistency term

MATLAB Routines

\( d, \, \text{red} \) Space dimension, number of uniform refinements
\( c4n, n4e \) Lists of coordinates of nodes, and nodes of elements
\( \text{Db}, \, \text{Nb} \) Lists of sides on \( \Gamma_D \) and \( \Gamma_N \)
\( \text{dNodes}, \, \text{fNodes} \) Nodes belonging to \( \Gamma_D \) and remaining nodes
\( nC, nE, \text{nDb}, \text{nNb} \) Number of nodes, elements and sides on \( \Gamma_D \) and \( \Gamma_N \)
\( s, m, \text{m_lumped} \) \( P1 \) stiffness, mass, and lumped mass matrix
\( m_{\text{Nb}}, \text{m_Nb_lumped} \) Exact and discrete inner products on \( \Gamma_N \)
\( \text{vol}_T, \text{mp}_T \) Volume and midpoint of an element
\( \text{grads}_T \) Elementwise gradients of nodal basis functions
\( \tau \) Step-size
\( I, J, X \) Lists to generate a sparse matrix

Special Constants

\( c_P \) Poincaré inequality
\( c_S \) Sobolev inequality
\( c_{\text{Tr}} \) Trace inequality
\( c_\Delta \) \( H^2 \)-regularity for \( \Delta \)
\( c_I \) Nodal interpolation
\( c_J \) Clément quasi-interpolation
\( c_{\text{usr}} \) Uniform shape regularity
\( c_{\text{inv}} \) Inverse estimate
\( \alpha, \beta, \gamma \) Ellipticity, inf-sup condition
\( k_a, k_b, k_I \) Continuity constants
C.2 Important Inequalities

Young’s Inequality For all \( a, b \in \mathbb{R}_{\geq 0} \) and \( 1 < p, q < \infty \) with \( 1/p + 1/q = 1 \) we have that
\[
ab \leq \frac{1}{p} a^p + \frac{1}{q} b^q.
\]

Cauchy–Schwarz Inequality For all \( x, y \in \mathbb{R}^n \) we have that
\[
x \cdot y = \sum_{i=1}^{n} x_i y_i \leq \left( \sum_{i=1}^{n} x_i^2 \right)^{1/2} \left( \sum_{i=1}^{n} y_i^2 \right)^{1/2} = |x||y|.
\]

Triangle Inequality For all \( x, y \in \mathbb{R}^n \) we have that
\[
|x + y| \leq |x| + |y|.
\]

Hölder’s Inequality For \( f \in L^p(\Omega) \) and \( g \in L^q(\Omega) \) and \( 1 \leq p, q \leq \infty \) with \( 1/p + 1/q = 1 \) we have that
\[
\int_{\Omega} fg \, dx \leq \|f\|_{L^p(\Omega)} \|g\|_{L^q(\Omega)}.
\]

Minkowski’s Inequality For \( f, g \in L^p(\Omega) \) with \( 1 \leq p \leq \infty \) we have that
\[
\|f + g\|_{L^p(\Omega)} \leq \|f\|_{L^p(\Omega)} + \|g\|_{L^p(\Omega)}.
\]

Poincaré (or Friedrichs) Inequality For \( v \in W^{1,p}(\Omega) \) with \( 1 \leq p \leq \infty \) and \( v|_{\partial \Omega} = 0 \) or \( \int_{\Omega} v \, dx = 0 \) we have that
\[
\|v\|_{L^p(\Omega)} \leq c_p \|\nabla v\|_{L^p(\Omega)}.
\]

Trace Inequality For \( v \in W^{1,p}(\Omega) \) and \( 1 \leq p \leq \infty \) we have that
\[
\|v\|_{L^p(\partial \Omega)} \leq c_{Tr} \|v\|_{W^{1,p}(\Omega)}.
\]
**Sobolev Inequalities** For $v \in W^{1,p}(\Omega)$ with $1 \leq p < d$ and $1 \leq q \leq p^*$ for $p^* = dp/(d-p)$ we have that

$$\|v\|_{L^q(\Omega)} \leq c_S\|v\|_{W^{1,p}(\Omega)}.$$

**Nodal Interpolation Estimate** If $v \in W^{2,p}(\Omega)$ with $p \geq 2$ we have that

$$\|v - \mathcal{I}_h v\|_{L^p(\Omega)} + h\|\nabla (v - \mathcal{I}_h v)\|_{L^p(\Omega)} \leq c\mathcal{I} h^2\|D^2 v\|_{L^p(\Omega)}.$$

**Uniform Shape Regularity** For all $h > 0$ and all $T \in \mathcal{T}_h$ we have that

$$h_T \leq c_{\text{ust}}q_T.$$

**Inverse Estimate** For all $v_h \in \mathcal{P}^1(\mathcal{T}_h)$ and $1 \leq p \leq \infty$ we have that

$$\|\nabla v_h\|_{L^p(\Omega)} \leq c_{\text{inv}}h^{-1}_{\text{min}}\|v_h\|_{L^p(\Omega)}.$$

**Discrete Norm Equivalence** For all $v_h \in \mathcal{P}^1(\mathcal{T}_h)$ we have that

$$c_{\text{eq}}^{-1}\|v_h\|_{L^p(\Omega)} \leq \left( \sum_{z \in \mathcal{T}_h} h_z^d |v_h(z)|^p \right)^{1/p} \leq c_{\text{eq}}\|v_h\|_{L^p(\Omega)}.$$

**$H^2$-Regularity** If $\Omega \subset \mathbb{R}^d$ is convex, then for all $v \in H^2(\Omega) \cap H_0^1(\Omega)$ we have that

$$\|D^2 v\|_{L^2(\Omega)} \leq c_\Delta \|\Delta v\|_{L^2(\Omega)}.$$

### C.3 Guidelines for Discretizing Differential Equations

The following aspects should be taken into account when developing a numerical scheme for approximating the solution of a partial differential equation.

1. Is the problem under consideration *well-posed* and what are typical properties of solutions, e.g., occurrence of characteristics, validity of a maximum principle, regularity properties, energy conservation principles?
2. Is the discretization *consistent*, i.e., does an interpolant of the exact solution nearly satisfy the numerical scheme?
3. Is the scheme stable, i.e., do numerical solutions remain bounded in a meaningful norm as discretization parameters tend to zero, e.g., is a discrete maximum principle or energy conservation principle satisfied?

4. Can the numerical solution be efficiently computed and is the discrete problem well-conditioned? What is a good stopping criterion for iterative methods? It is necessary for the problem to have the same number of degrees of freedom as conditions that characterize solutions.

5. Do numerical solutions reflect the typical behavior of the physical process with a reasonable discretization fineness, e.g., transport along characteristics?

6. Which terms should be discretized implicitly and which ones explicitly? Low-order terms, which involve lower-order derivatives, can often be treated explicitly which simplifies computations, while the explicit discretization of highest order derivatives typically leads to restrictive stability conditions.

7. If the discretization is not well-posed, stabilizing terms such as $-\varepsilon \Delta u$ can be introduced. This requires relating $\varepsilon$ to the mesh-size and typically leads to reduced convergence rates.

8. How do nonconformities such as implementation aspects affect approximations, e.g., the use of quadrature or domain approximation?

9. Does an appropriate reformulation of the problem allow for more stable and accurate schemes, e.g., integration is more stable than differentiation.

10. Has the implementation been debugged, e.g., by solving a simple discretization by hand, checking enumerations of sides and vertices, and testing the correctness of matrices.
Index

C¹-element, 110
H²-regularity, 94, 526
P₀-element, 100
P₁-element, 100, 113
P₂-element, 100
Q₀-element, 101
θ-method, 26
dG-norm, 336
h-h/2-estimator, 453

adaptative algorithm, 175
additive iteration, 230
adjoint operator, 259
affine family, 110
almost everywhere, 77
alternating Schwarz method, 228
Ampère's law, 378
ansatz functions, 99
approximation error, 132
approximation property, 213
Argyris element, 371
Arrow–Hurwicz algorithm, 258
asymptotical exactness, 178
Aubin–Nitsche lemma, 115
average, 333
averaging estimator, 454
a posteriori error estimate, 169

Babuška paradox, 370
Babuška–Brezzi conditions, 273
backward Euler scheme, 22
best approximation, 93, 114
biharmonic problem, 369
bilaplacian, 368
bisection, 176
blue-refinement, 176
bottleneck, 511
boundary approximation, 121
boundary condition, 4, 56
boundary layer, 319
boundary nodes, 54
boundary value problem, 46, 56
bounded bilinear form, 75
BPX preconditioner, 238
Bramble–Hilbert lemma, 105
Brezzi’s theorem, 265
broken Sobolev space, 309, 332
bubble function, 172, 304
bulk modulus, 350
Cauchy–Schwarz inequality, 70, 237, 525
Céa’s lemma, 92, 115, 272
cell values, 515
CFL condition, 12
characteristics, 5
Chorin projection scheme, 397
Clément quasi-interpolant, 166
clampered plate, 366
classical solution, 65
closed range theorem, 259
coarse-grid correction, 217
coarsening, 201
coercive bilinear form, 75
compact operator, 77
compatibility condition, 396
condition number, 249
cone condition, 91

© Springer International Publishing Switzerland 2016
S. Bartels, Numerical Approximation of Partial Differential Equations,
Texts in Applied Mathematics 64, DOI 10.1007/978-3-319-32354-1
conforming method, 304
conforming triangulation, 109
conjugate gradient algorithm, 232
consistency, 59, 323
corner singularity, 155
Crank–Nicolson scheme, 30
Crouzeix–Raviart element, 310
Cubic Hermite element, 101
curl operator, 380
d'Alembert's formula, 36
damping parameter, 221
de Rham complex, 381
deflection, 364
definition, 349
degrees of freedom, 134
Delaunay triangulation, 329
density, 85, 87
diameter, 106
difference quotient, 6, 7
diffusion equation, 16
Dirichlet boundary condition, 16, 46
Dirichlet energy, 69
Dirichlet–Neumann method, 221
discrete inner product, 143
discrete Laplacian, 203
displacement, 349
divergence theorem, 79
do-nothing condition, 392
Donald diagram, 331
dual formulation, 285
dual norm, 249
dual space, 73
Duffy's transformation, 119
Eberlein–Šmulian theorem, 77
edge bisection, 176
efficiency, 173
elastic material, 349
element, 99, 109
element patch, 172
elementwise gradient, 332
elliptic equation, 58
elliptic reconstruction, 203
enumeration, 52
error estimator, 170
Euler scheme, 20
Eulerian description, 391
experimental convergence rate, 133
explicit scheme, 8, 20
Faraday's law, 378
Fatou's lemma, 79
Fick's law, 15
finite element, 99
finite overlap, 164
finite volume method, 329
floating domain, 223
flux, 327
form, 68
formal equivalence, 68
Fortin interpolant, 276
Fourier symbol, 14
Fourier-stability, 13
frame indifference, 350
free nodes, 125
Friedrichs inequality, 88
fundamental lemma, 67, 83
Galerkin approximation, 92
Galerkin orthogonality, 93, 114
Gauss's law, 378
Gauss's theorem, 79
generalized eigenvalue, 212
ghost points, 33
global degrees of freedom, 110
global node numbers, 126
graded grid, 156
graduation strength, 156
Green's identity, 47
green-refinement, 176
grid function, 49
grid points, 5
Hahn–Banach theorem, 74
hanging node, 175
harmonic extension, 224
harmonic fields, 381
harmonic function, 46
heat equation, 16
Hellinger–Reissner principle, 354
Helmholtz decomposition, 381
Hilbert space, 70
Hölder's inequality, 78, 525
Hsiegh–Clough–Tocher element, 373
hyperbolic equation, 58
ill-posed problem, 57
implicit scheme, 23
Index

incompressible, 353, 392
inf-sup condition, 260
inflow boundary, 4, 320, 392
initial boundary value problem, 7
initial condition, 4
inner radius, 106
interface, 220
interior hot spots, 19
interior nodes, 54
interior penalty method, 336
interpolant, 102, 110, 114, 199
interpolation error, 123, 209, 526
isomorphism, 260
isoparametric element, 112
isotropy, 350

jump, 169, 302, 333

kinematic viscosity, 393
Kirchhoff bending energy, 365
Korn inequality, 351

Lagrange functional, 257
Lagrange multiplier, 257
Lagrangian description, 391
Lamé constants, 350
Laplace equation, 46
Laplace operator, 46, 56
Lax–Milgram lemma, 75, 262
Lax–Richtmyer theorem, 60
Lebesge measure, 77
Lebesgue integral, 78
Lebesgue space, 78
linear equation, 56
linear functional, 73
linear operator, 73
linear problem, 56
linearized strain, 349
load vector, 114
local conservation property, 328
local degrees of freedom, 101
local node numbers, 126
locking effect, 277
logarithmic slope, 133

mass lumping, 143
maximum angle condition, 111
maximum principle, 19, 117
Maxwell inequality, 381
Maxwell system, 377
mean value property, 47, 50
measurable function, 77
measurable set, 77
membrane phenomena, 366
Meyers–Serrin theorem, 85
midpoint scheme, 26
min-max problem, 257
MINI element, 304
minimum angle condition, 111
Minkowski’s inequality, 78, 525
monotonicity, 116
Morrey’s theorem, 90
multi-index, 84
multigrid algorithm, 216
multilevel preconditioner, 233
multiplicative iteration, 230

Navier–Lamé equations, 351
Navier–Stokes equations, 247, 392
Nédélec element, 385
nested spaces, 182
Neumann boundary condition, 16, 46
Newtonian fluid, 391
no-slip condition, 392
nodal basis, 113
nodal interpolant, 102
nodal interpolation, 526
node functional, 99
node patch, 162
nodes, 112
norm convergent, 76
norm equivalence, 209, 224, 526
null set, 77

Ohm’s law, 379
open mapping theorem, 74
operator norm, 73
order of an equation, 56
Oseen system, 395
outflow boundary, 320
overlap region, 227

parabolic boundary, 18, 320
parabolic equation, 58
partial differential equation, 4, 55
Péclét number, 319, 326
perfectly conducting, 379
Picard iteration, 395
Piola transformation, 290
plane strain, 351
plane stress, 351
plate, 364
plate bending, 369
Poincaré inequality, 90, 104, 525
point values, 515
Poisson problem, 46, 65, 91
polar set, 259
polynomials, 99
Prager–Synge theorem, 177
preasymptotic range, 134
preconditioner, 231
pressure stabilization, 306
primal form, 268
projection step, 396
prolongation, 129, 216

quasi-interpolant, 166
quasioptimal, 132
quasiuniform, 123

Raviart–Thomas element, 287
red-refinement, 176
refinement, 182
refinement edge, 176
refinement indicator, 169
reflexive space, 76
regular triangulations, 111
reliability, 170
Rellich–Kandrachov theorem, 90
residual, 172
restriction operator, 216
Reynold’s number, 395
Riesz representation, 72
rigid body motion, 349

saddle-point formulation, 382
saddle-point problem, 257
Schur complement, 258
Scott–Zhang quasi-interpolant, 185
separable space, 77
separating hyperplane, 74
separation of variables, 18, 408
shape regular, 111
shear modulus, 350
side patch, 172
sides, 165
simple support, 366
singular perturbation, 320
singularity function, 156

smoothing property, 213
Sobolev exponent, 89
Sobolev function, 82
Sobolev inequality, 89, 526
Sobolev space, 84
softening effect, 279
solution, 56
stability, 10, 59
stationary, 46
stencil, 49
step-size, 5
stiffness matrix, 93, 114, 209, 505
Stokes system, 247
Stokes’s theorem, 378
strain tensor, 349
Strang lemma, 118, 123
streamline-diffusion method, 323
stress tensor, 350
strongly convergent, 76
subcharacteristic, 320
superconvergence, 134
surface force, 350
symmetric gradient, 349

Taylor–Hood element, 305
time horizon, 7
tow-level preconditioner, 243
trace inequality, 525
transformation, 79, 106
transmission condition, 220
transport equation, 4
triangle inequality, 525
triangulation, 109

unconditionally stable, 25, 59
uniform shape regularity, 526
upwinding, 13, 331
Uzawa algorithm, 258

volume force, 350
von Neumann-stability, 13
Voronoi polygons, 329

wave equation, 35
weak form, 300
weak formulation, 67, 381, 394
weak gradient, 82
<table>
<thead>
<tr>
<th>Term</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>weak limit</td>
<td>76</td>
</tr>
<tr>
<td>weak solution</td>
<td>91, 140, 145, 197</td>
</tr>
<tr>
<td>weakly acute</td>
<td>117, 330</td>
</tr>
<tr>
<td>weakly convergent</td>
<td>76</td>
</tr>
<tr>
<td>weakly differentiable</td>
<td>82</td>
</tr>
<tr>
<td>well-posed problem</td>
<td>57</td>
</tr>
<tr>
<td>Wheeler’s trick</td>
<td>144</td>
</tr>
<tr>
<td>Young’s inequality</td>
<td>78, 525</td>
</tr>
</tbody>
</table>
Finite element methods for approximating partial differential equations have reached a high degree of maturity, and are an indispensable tool in science and technology. This textbook aims at providing a thorough introduction to the construction, analysis, and implementation of finite element methods for model problems arising in continuum mechanics. The first part of the book discusses elementary properties of linear partial differential equations along with their basic numerical approximation, the functional-analytical framework for rigorously establishing existence of solutions, and the construction and analysis of basic finite element methods. The second part is devoted to the optimal adaptive approximation of singularities and the fast iterative solution of linear systems of equations arising from finite element discretizations. In the third part, the mathematical framework for analyzing and discretizing saddle-point problems is formulated, corresponding finite element methods are analyzed, and particular applications including incompressible elasticity, thin elastic objects, electromagnetism, and fluid mechanics are addressed. The book includes theoretical problems and practical projects for all chapters, an introduction to the implementation of finite element methods, and model implementations of most devised schemes.