Lecture Notes for the Course Numerical Methods for Continuum Mechanics 1

Walter Zulehner
Institute for Computational Mathematics
Johannes Kepler University Linz

Summer Semester 2011

Contents

Chapter 1

Models

1.1 Kinematics

Let $\Omega \subset \mathbb{R}^3$ be an open, bounded and connected set with Lipschitz-continuous boundary $\Gamma = \partial\Omega$. The set $\overline{\Omega}$ is called the reference configuration and describes, e.g., the initial state or the undeformed state of a continuum (body).

A configuration (or deformation) is a sufficiently smooth, orientation preserving and injective mapping

$$\phi \colon \overline{\Omega} \longrightarrow \mathbb{R}^3.$$

This mapping describes, e.g., the state of the continuum at some later time or the state of a deformed continuum. The set $\phi(\overline{\Omega})$ consists of all points (or particles) x of the form

$$x = \phi(X)$$

with $X \in \overline{\Omega}$. X are called the material (or Lagrangian) coordinates, x are called the spatial (or Eulerian) coordinates of a particle.

The (Jacobian) matrix

$$\mathbf{F}(X) = \nabla\phi(X) = \begin{pmatrix} \frac{\partial\phi_1}{\partial X_1}(X) & \frac{\partial\phi_1}{\partial X_2}(X) & \frac{\partial\phi_1}{\partial X_3}(X) \\ \frac{\partial\phi_2}{\partial X_1}(X) & \frac{\partial\phi_2}{\partial X_2}(X) & \frac{\partial\phi_2}{\partial X_3}(X) \\ \frac{\partial\phi_3}{\partial X_1}(X) & \frac{\partial\phi_3}{\partial X_2}(X) & \frac{\partial\phi_3}{\partial X_3}(X) \end{pmatrix}$$

is called the deformation gradient. Preserving the orientation corresponds to the condition

$$J(X) = \det \nabla\phi(X) > 0 \quad \text{for all } X \in \overline{\Omega}.$$

The displacement $U \colon \overline{\Omega} \longrightarrow \mathbb{R}^d$, introduced by

$$U(X) = x - X \quad \text{with} \quad x = \phi(X)$$

measures the deviation from the reference configuration. For

$$x = \phi(X), \quad \overline{x} = \phi(\overline{X}), \quad \Delta X = \overline{X} - X, \quad \Delta x = \overline{x} - x,$$

we have:

$$\Delta x = \phi(X + \Delta X) - \phi(X) = \nabla\phi(X)\Delta X + o(|\Delta X|)$$

with the notation

$$|a| = \sqrt{a^T a} = \sqrt{\sum_{i=1}^{d} a_i^2} \quad \text{for } a \in \mathbb{R}^d.$$

Therefore,

$$\begin{aligned} |\Delta x|^2 &= \Delta X^T \nabla\phi(X)^T \nabla\phi(X)\Delta X + o(|\Delta X|^2) \\ &= \Delta X^T \mathbf{C}(x)\Delta X + o(|\Delta X|^2) \end{aligned}$$

with

$$\mathbf{C}(X) = \mathbf{F}(X)^T\mathbf{F}(X) = \nabla\phi(X)^T\nabla\phi(X).$$

The symmetric tensor $\mathbf{C}(X)$ is called the (right) Cauchy-Green deformation tensor. It describes the local change in distances by the deformation.

Remark: $\mathbf{C}(X)$ also describes the local change in angles: For

$$x = \phi(X), \quad x_i = \phi(X_i), \quad \Delta X_i = X_i - X, \quad \Delta x_i = x_i - x \quad \text{for} \quad i = 1, 2$$

we have:

$$\Delta x_i = \phi(X + \Delta X_i) - \phi(X) = \nabla\phi(X)\Delta X_i + o(\Delta X_i).$$

Therefore

$$\begin{aligned} \Delta x_1 \cdot \Delta x_2 = \Delta x_2^T \Delta x_1 &= \Delta X_2^T \nabla\phi(X)^T \nabla\phi(X)\Delta X_1 + o(|\Delta X_1|\,|\Delta X_2|) \\ &= \Delta X_2^T \mathbf{C}(x)\Delta X_1 + o(|\Delta X_1|\,|\Delta X_2|) \end{aligned}$$

with the notation

$$a \cdot b = b^T a = \sum_{i=1}^{d} a_i b_i \quad \text{for } a, b \in \mathbb{R}^d.$$

It can be shown that there is no change in distances, i.e.:

$$\mathbf{C}(X) = I \quad \text{for all } X \in \Omega,$$

if and only if the configuration is a rigid body configuration, i.e.:

$$\phi(X) = QX + a,$$

where Q is an orthogonal matrix with $\det Q = 1$ (describing a rotation) and $a \in \mathbb{R}^3$ (describing a translation).

The deviation of $\mathbf{C}(X)$ from the ideal case I is measured by the symmetric tensor

$$\mathbf{E}(X) = \frac{1}{2}(\mathbf{C}(X) - I),$$

the so called Green-St.Venant strain tensor. Then, of course, we have:

$$|\Delta x|^2 - |\Delta X|^2 = 2\,\Delta X^T \mathbf{E}(X) \Delta X + o(|\Delta X|^2).$$

$\mathbf{E}(X)$ can be expressed directly by the displacement $U(X)$:

$$\mathbf{E}[U](X) = \frac{1}{2}\left(\nabla U(X)^T + \nabla U(X) + \nabla U(X)^T \nabla U(X)\right),$$

or, component-wise:

$$E_{ij}[U](X) = \frac{1}{2}\left(\frac{\partial U_j}{\partial X_i}(X) + \frac{\partial U_i}{\partial X_j}(X) + \sum_k \frac{\partial U_k}{\partial X_i}(X)\frac{\partial U_k}{\partial X_j}(X)\right).$$

Observe the nonlinear relation between $\mathbf{E}$ and U.

The displacement can also be introduced in Eulerian coordinates by

$$u(x) = x - X \quad \text{with} \quad x = \phi(X), \text{ i.e. } u(x) = x - \phi^{-1}(x).$$

For

$$X = \phi^{-1}(x), \quad \overline{X} = \phi^{-1}(\overline{x}), \quad \Delta X = \overline{X} - X, \quad \Delta x = \overline{x} - x,$$

we have:

$$\begin{aligned}\Delta X &= \phi^{-1}(x + \Delta x) - \phi^{-1}(x) = \nabla\left(\phi^{-1}\right)(x)\Delta x + o(\Delta x) \\ &= (\nabla\phi(X))^{-1}\Delta x + o(\Delta x) \quad \text{with } X = \phi^{-1}(x)\end{aligned}$$

and, consequently,

$$|\Delta X|^2 = \Delta x^T \mathbf{c}(x) \Delta x + o(|\Delta x|^2),$$

where

$$\mathbf{c}(x) = \mathbf{b}(x)^{-1} \quad \text{with} \quad \mathbf{b}(x) = \mathbf{F}(X)\mathbf{F}(X)^T = \nabla\phi(X)\nabla\phi(X)^T \quad \text{for } X = \phi^{-1}(x).$$

Furthermore,

$$|\Delta x|^2 - |\Delta X|^2 = 2\,\Delta x^T \mathbf{e}(x) \Delta x + o(|\Delta x|^2).$$

with

$$\mathbf{e}(x) = \frac{1}{2}(I - \mathbf{c}(x)).$$

Finally, it easily follows that

$$\mathbf{e}[u](x) = \frac{1}{2}\left(\nabla u(x)^T + \nabla u(x) - \nabla u(x)^T \nabla u(x)\right).$$

$\mathbf{b}(x)$ is called the Finger deformation tensor or the left Cauchy-Green deformation tensor, $\mathbf{e}(x)$ is called the Almansi-Hamel strain tensor or the Euler strain tensor.

The motion of a continuum (or body) is described by a curve

$$t \mapsto \phi_t.$$

Interpretation: The position x of a point (particle) at time t, whose position at time 0 was X, is given by

$$x = \phi_t(X) \equiv \phi(X, t).$$

Then the material (or Lagrangian) velocity of this particle as a function of X and t is given by

$$V_t(X) = V(X, t) = \frac{\partial \phi}{\partial t}(X, t),$$

and the material (or Lagrangian) acceleration is given by

$$A_t(X) = A(X, t) = \frac{\partial^2 \phi}{\partial t^2}(X, t).$$

Observe the following linear relation between velocity and acceleration:

$$A(X, t) = \frac{\partial V}{\partial t}(X, t).$$

In the Eulerian approach the motion of a particle is described by the spatial velocity (field) $v(x, t)$, where $v(x, t)$ is the velocity of that particle, which passes through x at time t, so

$$v_t(x) = v(x, t) = V(X, t) = \frac{\partial \phi}{\partial t}(X, t) \text{ with } x = \phi(X, t),$$

i.e.:

$$v(x, t) = \frac{\partial \phi}{\partial t}(\phi_t^{-1}(x), t).$$

For the spatial acceleration $a(x, t)$ of that particle we obtain:

$$a_t(x) = a(x, t) = A(X, t) = \frac{\partial^2 \phi}{\partial t^2}(X, t) \text{ with } x = \phi(X, t).$$

We have for $x = \phi(X, t)$:

$$a(x, t) = \frac{\partial}{\partial t}[v(\phi(X, t), t)] = \frac{\partial v}{\partial t}(x, t) + \sum_i v_i(x, t) \frac{\partial v}{\partial x_i}(x, t).$$

Notation: The differential operator $v \cdot \text{grad} = v \cdot \nabla$, given by

$$(v \cdot \text{grad}) f = (v \cdot \nabla) f = \sum_{i=1}^{d} v_i \frac{\partial f}{\partial x_i},$$

is called the convective derivative and the differential operator d/dt, given by

$$\frac{df}{dt} = \dot{f} = \frac{\partial f}{\partial t} + (v \cdot \text{grad}) f,$$

is called the total or material derivative.

With these notations the spatial acceleration can be written in the following form:

$$a(x,t) = \frac{dv}{dt}(x,t) = \frac{\partial v}{\partial t}(x,t) + (v(x,t) \cdot \text{grad}) v(x,t) = \frac{\partial v}{\partial t}(x,t) + (v(x,t) \cdot \nabla) v(x,t).$$

Observe that this is a nonlinear relation between velocity and acceleration in the Eulerian approach.

Remark: For a given velocity (field) $v(x,t)$ one obtains the trajectories $\phi(X,t)$ of the individual particles as solution of the initial value problem:

$$\begin{aligned} \frac{\partial \phi}{\partial t}(X,t) &= v(\phi(X,t),t), \\ \phi(X,0) &= X. \end{aligned}$$

1.2 Balance Laws

The set

$$\overline{\Omega}_t = \phi_t(\overline{\Omega}) = \{\phi(X,t) \mid X \in \overline{\Omega}\}$$

describes the position of all particles from the reference configuration at time t. Let $\omega \subset \overline{\Omega}$ be an open set with Lipschitz-continuous boundary. Then the set ω_t, given by

$$\omega_t = \phi_t(\omega) = \{\phi(X,t) \mid X \in \omega\},$$

describes the position of those particles at time t, which were in ω at time $t = 0$.

1.2.1 The Reynolds Transport Theorem

The Reynolds transport theorem describes the rate change of the quantity

$$\mathcal{F}(t) = \int_{\omega_t} F(x,t)\, dx$$

for a given function F of x and t:

Theorem 1.1 (Reynolds transport theorem)**.** *Let ϕ be twice continuously differentiable and F continuously differentiable. Then*

$$\frac{d\mathcal{F}}{dt}(t) = \int_{\omega_t} \left[\frac{\partial F}{\partial t}(x,t) + \operatorname{div}(Fv)(x,t) \right] \, dx = \int_{\omega_t} \left[\frac{dF}{dt}(x,t) + F \, \operatorname{div}(v)(x,t) \right] \, dx.$$

Notation: $\operatorname{div} G = \nabla \cdot G$, given by

$$\operatorname{div} G = \nabla \cdot G = \sum_{i=1}^{d} \frac{\partial G_i}{\partial x_i}$$

for a continuously differentiable vector-valued function G, is called the divergence of G.

Remark: With the help of Gauss' theorem it follows immediately that

$$\frac{d\mathcal{F}}{dt}(t) = \int_{\omega_t} \frac{\partial F}{\partial t} \, dx + \int_{\partial\omega_t} F \; v \cdot n \; ds.$$

Here $n = n(x,t)$ denotes the outer normal unit vector at a point x on the boundary of ω_t.

1.2.2 Conservation of Mass

Let $\rho(x,t)$ denote the mass density of a body at the position x and time t. The principle of conservation of mass states that no mass will be generated or destroyed, i. e.:

$$\frac{d}{dt} \int_{\omega_t} \rho(x,t) \, dx = 0.$$

Under appropriate smoothness conditions the transport theorem implies:

$$\int_{\omega_t} \left[\frac{\partial \rho}{\partial t}(x,t) + \operatorname{div}(\rho v)(x,t) \right] \, dx = 0$$

for all t and all open sets $\omega \subset \overline{\Omega}$ with Lipschitz-continuous boundary. This results in the following differential equation, the so-called equation of continuity: either in conservative form:

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho v) = 0,$$

or, equivalently, in the convective form:

$$\frac{d\rho}{dt} + \rho \, \operatorname{div} v = 0.$$

In the special case $\rho =$ constant (incompressible fluid) the equation of continuity is given by

$$\operatorname{div} v = 0. \tag{1.1}$$

We have (by the substitution rule)

$$\int_{\omega_t} \rho(x,t)\ dx = \int_{\omega} \rho(\phi(X,t))J(X,t)\ dX.$$

Hence, the conservation of mass in Lagrangian coordinates reads:

$$\frac{d}{dt}\left(\rho(\phi(X,t),t)J(X,t)\right) = 0,$$

Therefore,

$$\rho(x,t) = \frac{1}{J(X,t)}\,\rho_0(X) \quad \text{with } x = \phi(X,t) \quad \text{and } \rho_0(X) = \rho(X,0).$$

1.2.3 Balance of Momentum and Angular Momentum

The total (linear) momentum of all particles in ω_t is given by

$$\int_{\omega_t} \rho(x,t)v(x,t)\ dx.$$

Newton's second law states that the rate of change of the (linear) momentum is equal to the applied forces $F(\omega_t)$, hence

$$\frac{d}{dt}\int_{\omega_t} \rho(x,t)v(x,t)\ dx = F(\omega_t).$$

The forces acting on the body can be split into applied body forces $F_V(\omega_t)$ and applied surface forces $F_S(\omega_t)$:

$$F(\omega_t) = F_V(\omega_t) + F_S(\omega_t).$$

If the body forces can be described by a specific force density (force per unit mass) $f(x,t)$, then we obtain the representation

$$F_V(\omega_t) = \int_{\omega_t} \rho(x,t)f(x,t)\ dx.$$

An example of such a force is the force of gravity with $f = (0,0,-g)^T$.

The internal surface forces can be described by a vector $\vec{t}(x,t,n)$ (force per unit area), the so-called Cauchy stress vector:

$$F_S(\omega_t) = \int_{\partial\omega_t} \vec{t}(x,t,n(x,t))\ ds.$$

Summarizing, we obtain the following balance law for the momentum:

$$\frac{d}{dt}\int_{\omega_t} \rho(x,t)v(x,t)\ dx = \int_{\omega_t} \rho(x,t)f(x,t)\ dx + \int_{\partial\omega_t} \vec{t}(x,t,n(x,t))\ ds.$$

The total angular momentum of all particles in ω_t is given by

$$\int_{\omega_t} x \times \rho(x,t)v(x,t)\ dx.$$

Newton's second law states that the rate of change of the angular momentum is equal to the applied torque, so

$$\frac{d}{dt}\int_{\omega_t} x \times \rho(x,t)v(x,t)\ dx = \int_{\omega_t} x \times \rho(x,t)f(x,t)\ dx + \int_{\partial\omega_t} x \times \vec{t}(x,t,n(x,t))\ ds.$$

These two equations are also called equations of motion, in the steady state case, also the equilibrium conditions.

Under reasonable assumptions it can be shown that the stress vector $\vec{t}(x,t,n) = (t_i(x,t,n))_{i=1,2,3}$ can be represented by the so-called Cauchy stress tensor $\sigma = (\sigma_{ij})$ in the following form:

$$t_i(x,t,n) = \sum_j \sigma_{ji}(x,t)\, n_j.$$

Using Gauss' theorem and the transport theorem one obtains for sufficiently smooth functions the following differential equation (in conservative form):

$$\frac{\partial}{\partial t}(\rho v_i) + \operatorname{div}(\rho v_i v) = \sum_j \frac{\partial \sigma_{ji}}{\partial x_j} + \rho f_i$$

from the balance of momentum, or in convective form

$$\rho\frac{\partial v_i}{\partial t} + \rho v \cdot \operatorname{grad} v_i = \sum_j \frac{\partial \sigma_{ji}}{\partial x_j} + \rho f_i$$

by using the equation of continuity,

It can be shown that the balance of angular momentum is satisfied if and only if σ is symmetric:

$$\sigma^T = \sigma.$$

Therefore, the balance of momentum in convective form can also be written in the following form:

$$\rho\frac{\partial v}{\partial t} + \rho(v \cdot \operatorname{grad})v = \operatorname{div}\sigma + \rho f$$

with

$$\operatorname{div}\sigma = \left(\sum_j \frac{\partial \sigma_{ij}}{\partial x_j}\right)_{i=1,2,3}.$$

So far, the equations of motion have been derived in Eulerian coordinates.

By transforming the integrals one easily obtains the equations of motion in Lagrangian coordinates. We have:

$$\int_{\omega_t} \rho(x,t)v(x,t)\ dx = \int_{\omega} \rho_0(X)V(X,t)\ dX$$
$$\int_{\omega_t} \rho(x,t)f(x,t)\ dx = \int_{\omega} \rho_0(X)F(X,t)\ dX$$
$$\int_{\partial\omega_t} \sigma(x,t)n(x,t)\ ds = \int_{\partial\omega} \mathbf{P}(X,t)N(X)\ dS$$

with the specific force density $F(X,t)$ in Lagrangian coordinates:

$$F(X,t) = f(x,t) \quad \text{for } x = \phi(X,t),$$

the unit normal vector $N(X)$ in Lagrangian coordinates:

$$\nabla\phi(X,t)^{-T}N(X) = |\nabla\phi(X,t)^{-T}N(X)|\, n(x,t) \quad \text{for } x = \phi(X,t),$$

and

$$\mathbf{P}(X,t) = J(X,t)\,\sigma(x,t)\nabla\phi(X,t)^{-T} \quad \text{for } x = \phi(X,t),$$

the so-called first Piola Kirchhoff stress tensor.

Remark: The last transformation rule is based on Nanson's formula:

$$\int_{\partial\omega_t} \sigma(x,t)n(x,t)\ ds = \int_{\partial\omega} \sigma(x,t)\, J(X,t)\, \nabla\phi(X,t)^{-T}N(X)\ dS.$$

Then one obtains from the balance of momentum the following differential equation in Lagrangian coordinates:

$$\rho_0(X)\frac{\partial^2\phi}{\partial t^2}(X,t) - \operatorname{div}\mathbf{P}(X,t) = \rho_0(X)F(X,t).$$

The balance of angular momentum is satisfied if and only if

$$\mathbf{S}(X,t)^T = \mathbf{S}(X,t)$$

with

$$\mathbf{S}(X,t) = \nabla\phi(X,t)^{-1}\mathbf{P}(X,t) = J(X,t)\,\nabla\phi(X,t)^{-1}\sigma(x,t)\nabla\phi(X,t)^{-T} \quad \text{for } x = \phi(X,t),$$

the so-called second Piola Kirchhoff stress tensor.

The corresponding transformation of the tensors $\mathbf{S} \mapsto \sigma$, given by

$$\sigma(x,t) = \frac{1}{J(X,t)}\nabla\phi(X,t)\mathbf{S}(X,t)\nabla\phi(X,t)^T \quad \text{for } x = \phi(X,t)$$

is called the Piola transformation.

Remark: Other balance laws like the balance of energy will not be discussed here.

1.3 Constitutive Laws

The equations of motion do not yet completely describe the configuration of a body. Equations for the stress in form of a constitutive laws are necessary.

Two important special cases will be considered here:

1.3.1 Elastic Materials

A material is called elastic if there is a constitutive law of the form

$$\mathbf{S}(X) = \hat{\mathbf{S}}(X, \mathbf{E}(X)).$$

For the important sub-class of hyperelastic materials the constitutive law can be represented by an energy functional:

$$\hat{\mathbf{S}}(X, \mathbf{E}) = \frac{\partial \Psi}{\partial \mathbf{E}}(X, \mathbf{E}),$$

where $\Psi(X, \mathbf{E})$ is the so-called stored energy function.

A material is called linearly elastic if

$$\Psi(X, \mathbf{E}) = \frac{1}{2} \sum_{ijkl} C_{ijkl}(X) E_{ij} E_{kl},$$

where the so-called elastic coefficients (or elasticity coefficients) $C_{ijkl}(X)$ (which form the so-called elasticity tensor) have the following properties:

$$C_{ijkl}(X) = C_{klij}(X)$$

and

$$C_{ijkl}(X) = C_{jikl}(X) = C_{jilk}(X).$$

From these conditions it follows that only 21 coefficients can be chosen independently from each other. For the corresponding constitutive law we obtain the linear relations:

$$S_{ij} = \sum_{kl} C_{ijkl}(X) E_{kl}, \tag{1.2}$$

which is called Hooke's law.

An important special case of linearly elastic materials are the St.Venant-Kirchhoff materials (homogenous, isotropic, and linearly elastic materials), for which the constitutive law has the form

$$\mathbf{S} = \lambda \, \mathrm{trace}(\mathbf{E}) \, I + 2\mu \, \mathbf{E}.$$

The parameters λ and μ are called Lamé coefficients. They are related to Young's modulus (or modulus of elasticity) E and Poisson's ratio ν by

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu}, \quad \nu = \frac{\lambda}{2(\lambda + \mu)}$$

and, vice versa

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \quad \mu = \frac{E}{2(1+\nu)}.$$

It can be shown by arguments from physics that:

$$0 < \nu < \frac{1}{2} \text{ and } E > 0.$$

These conditions are equivalent to

$$\lambda > 0 \text{ and } \mu > 0.$$

For St.Venant-Kirchhoff materials the stored energy function takes the form

$$\Psi(\mathbf{E}) = \frac{\lambda}{2}\,(\text{trace}(\mathbf{E}))^2 + \mu\ \text{trace}(\mathbf{E}^2),$$

so

$$C_{ijkl} = \lambda\,\delta_{ij}\,\delta_{kl} + \mu\,(\delta_{ik}\,\delta_{jl} + \delta_{il}\,\delta_{jk}).$$

1.3.2 Newtonian Fluids

Starting point is the following ansatz for the Cauchy stress tensor

$$\sigma = -p\,I + \tau,$$

where $p(x,t)$ denotes the pressure in the fluid at the position x and time t and τ depends on the first spatial derivative of the velocity field $v(x,t)$.

For a parallel flow (in x_1 direction) Newton postulated the linear relation

$$\tau_{21} = \mu\frac{dv_1}{dx_2}$$

for the shear stress τ_{21}. The coefficient μ is called the dynamic viscosity of the fluid.

Under reasonable assumptions it can be shown that this implies the following form for τ:

$$\tau = \lambda\ \text{div}\, v\ I + 2\mu\,\varepsilon(v)$$

with

$$\varepsilon(v) = (\varepsilon(v)_{ij}), \quad \varepsilon(v)_{ij} = \frac{1}{2}\left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i}\right).$$

Observe that $\text{div}\, v = \text{trace}\,\varepsilon(v)$ and the formal similarity to the constitutive law for St. Venant-Kirchhoff materials.

Arguments from physics show that

$$\mu \geq 0 \quad \text{and} \quad \hat{\mu} = \lambda + \frac{2}{3}\mu \geq 0.$$

The coefficient $\hat{\mu}$ is called bulk viscosity. In the following we will assume that $\hat{\mu} = 0$, hence $\lambda = -2\mu/3$. Therefore

$$\sigma = -(p + \frac{2\mu}{3} \operatorname{div} v)\, I + 2\mu\, \varepsilon(v).$$

For $\rho =$ constant, $\mu =$ constant and with the help of (1.1) ($\operatorname{div} v = 0$) the expressions for the internal surface force can be further simplified:

$$\operatorname{div} \sigma = -\operatorname{grad} p + \mu\, \Delta v,$$

where Δ denotes the Laplacian operator:

$$\Delta = \sum_{j=1}^{d} \frac{\partial^2}{\partial x_j^2}.$$

1.4 Boundary Value and Initial-Boundary Value Problems

For a complete description we need boundary conditions and for time-dependent problems initial conditions.

1.4.1 Elastostatics and Elastodynamics

Usually Lagrangian coordinates are used in elasticity.

In typical applications the surface force is prescribed on some part Γ_N of the boundary $\Gamma = \partial\Omega$ of Ω, given by its surface force density $T_N(x)$. This results in the boundary condition

$$(\nabla\phi\, \mathbf{S})\, N = T_N \quad \text{for all } x \in \Gamma_N,\ t > 0.$$

For the remaining part Γ_D of the boundary we assume that the deformation is known. This leads to the boundary condition

$$\phi = \phi_D \quad \text{for all } X \in \Gamma_D,\ t > 0.$$

As initial conditions usually the initial configuration and the initial velocity are prescribed:

$$\phi = \phi_0, \quad \frac{\partial \phi}{\partial t} = V_0 \quad \text{for } t = 0.$$

Hence we obtain the following initial-boundary value problem of elastodynamics:

$$
\begin{aligned}
\rho_0 \frac{\partial^2 \phi}{\partial t^2} - \operatorname{div}(\nabla\phi\, \mathbf{S}) &= \rho_0\, F && \text{in } \Omega,\ t > 0,\\
\mathbf{S} &= \hat{\mathbf{S}}(\mathbf{E}) && \text{in } \Omega,\ t > 0,\\
\mathbf{E} &= \frac{1}{2}(\nabla\phi^T \nabla\phi - I) && \text{in } \Omega,\ t > 0,\\
\phi &= \phi_D && \text{on } \Gamma_D,\ t > 0,\\
(\nabla\phi\, \mathbf{S})\, N &= T_N && \text{on } \Gamma_N,\ t > 0,\\
\phi = \phi_0, \quad \frac{\partial \phi}{\partial t} &= V_0 && \text{in } \Omega,\ t = 0.
\end{aligned}
$$

The corresponding time-independent problem leads to the following boundary value problem of elastostatics:

$$
\begin{aligned}
-\operatorname{div}(\nabla\phi\, \mathbf{S}) &= \rho_0\, F && \text{in } \Omega,\\
\mathbf{S} &= \hat{\mathbf{S}}(\mathbf{E}) && \text{in } \Omega,\\
\mathbf{E} &= \frac{1}{2}(\nabla\phi^T \nabla\phi - I) && \text{in } \Omega,\\
\phi &= \phi_D && \text{on } \Gamma_D,\\
(\nabla\phi\, \mathbf{S})\, N &= T_N && \text{on } \Gamma_N.
\end{aligned}
$$

1.4.2 Linear(ized) Elasticity

For small displacements it is justified

- not to distinguish between the Eulerian and the Lagrangian description (in the sequel we will use the Eulerian description), and
- to replace the strain tensor by the linearized strain tensor ε, given by

$$\varepsilon_{ij}(u) = \frac{1}{2}\left(\frac{\partial u_j}{\partial x_i} + \frac{\partial u_i}{\partial x_j}\right).$$

Then Hooke's law (1.2) can be written in the form

$$\sigma_{ij} = \sum_{kl} C_{ijkl}\, \varepsilon_{kl}$$

or, in short,

$$\sigma = C\,\varepsilon.$$

We obtain the following initial-boundary value problem of linear(ized) elastodynamics:

$$\begin{aligned}
\rho_0 \frac{\partial^2 u}{\partial t^2} - \operatorname{div}\sigma &= \rho_0 f && \text{in } \Omega,\ t > 0,\\
\sigma &= C\,\varepsilon && \text{in } \Omega,\ t > 0,\\
\varepsilon &= \frac{1}{2}(\nabla u^T + \nabla u) && \text{in } \Omega,\ t > 0,\\
u &= u_D && \text{on } \Gamma_D,\ t > 0,\\
\sigma\ n &= t_N && \text{on } \Gamma_N,\ t > 0,\\
u = u_0, \quad \frac{\partial u}{\partial t} &= v_0 && \text{in } \Omega,\ t = 0,
\end{aligned}$$

and the following boundary value problem of linear(ized) elastostatics:

$$\begin{aligned}
-\operatorname{div}\sigma &= \rho_0 f && \text{in } \Omega,\\
\sigma &= C\,\varepsilon && \text{in } \Omega,\\
\varepsilon &= \frac{1}{2}(\nabla u^T + \nabla u) && \text{in } \Omega,\\
u &= u_D && \text{on } \Gamma_D,\\
\sigma\ n &= t_N && \text{on } \Gamma_N.
\end{aligned}$$

For St. Venant-Kirchhoff materials we obtain, in particular,

$$\sigma = \lambda\ \operatorname{trace}(\varepsilon)\, I + 2\,\mu\,\varepsilon$$

and from constitutive law and the linearized strain-displacement relations it follows that:

$$\begin{aligned}
-\operatorname{div}\sigma &= -2\,\mu\ \operatorname{div}\varepsilon(u) - \lambda\ \operatorname{grad}\operatorname{div} u\\
&= -\mu\,\Delta u - (\lambda + \mu)\ \operatorname{grad}\operatorname{div} u.
\end{aligned}$$

The corresponding second order differential equations for the displacement u are called Lamé (or Cauchy-Navier) equations.

1.4.3 The Navier-Stokes Equations

Usually Eulerian coordinates are used in fluid mechanics. The unknown functions are, e.g., the velocity $v(x,t)$ and the pressure $p(x,t)$.

In typical applications the surface force is prescribed on some part Γ_N of the boundary $\Gamma = \partial\Omega$ of Ω, given by its surface force density $t_N(x)$. This results in the boundary condition

$$\sigma n = t_N \quad \text{for all } x \in \Gamma_N,\ t > 0.$$

For the remaining part Γ_D of the boundary we assume that the velocity is known. This leads to the boundary condition

$$v = v_D \quad \text{for all } x \in \Gamma_D,\ t > 0.$$

As initial condition usually the initial velocity is prescribed:

$$v = v_0 \quad \text{for } t = 0.$$

For the case ρ = constant and μ = constant one obtains the equations of motion in conservative form

$$\frac{\partial}{\partial t}(\rho v_i) + \operatorname{div}(\rho v_i v) = -\frac{\partial p}{\partial x_i} + \mu \Delta v_i + \rho f_i, \tag{1.3}$$

or in convective form

$$\rho \frac{\partial v}{\partial t} + \rho\,(v \cdot \operatorname{grad}) v = -\operatorname{grad} p + \mu\,\Delta v + \rho f \tag{1.4}$$

or, after dividing by ρ:

$$\frac{\partial v}{\partial t} + (v \cdot \operatorname{grad}) v = -\frac{1}{\rho} \operatorname{grad} p + \nu \Delta v + f \tag{1.5}$$

with $\nu = \mu/\rho$, the kinematic viscosity. The equations (1.3) or (1.4) or (1.5) are called the Navier-Stokes equations.

In summary, one obtains the following initial-boundary value problem of fluid mechanics:

$$\begin{aligned}
\frac{\partial v}{\partial t} + (v \cdot \operatorname{grad}) v - \nu\,\Delta v + \frac{1}{\rho} \operatorname{grad} p &= f \quad && \text{in } \Omega,\ t > 0,\\
\operatorname{div} v &= 0 \quad && \text{in } \Omega,\ t > 0,\\
v &= v_D \quad && \text{on } \Gamma_D,\ t > 0,\\
\sigma\ n &= t_N \quad && \text{on } \Gamma_N,\ t > 0,\\
v &= v_0 \quad && \text{in } \Omega,\ t = 0,
\end{aligned}$$

and, for the steady state case, the corresponding boundary value problem:

$$\begin{aligned}(v \cdot \operatorname{grad})v - \nu\,\Delta v + \frac{1}{\rho}\operatorname{grad} p &= f \quad && \text{in } \Omega,\\ \operatorname{div} v &= 0 && \text{in } \Omega,\\ v &= v_D && \text{on } \Gamma_D,\\ \sigma\ n &= t_N && \text{on } \Gamma_N.\end{aligned}$$

Dimensional analysis:

Starting from reference values L^*, t^*, U^* and p^* for the length, the time, the velocity and the pressure new variables are introduced by

$$x_i' = \frac{x_i}{L^*},\ t_i' = \frac{t}{t^*},\ v_i' = \frac{v_i}{U^*},\ p' = \frac{p}{p^*}.$$

By transformation of variables one obtains:

$$\frac{\rho U^*}{t^*}\frac{\partial v_i'}{\partial t'} + \frac{\rho (U^*)^2}{L^*}\sum_j v_j'\frac{\partial v_i'}{x_j'} - \frac{\mu U^*}{(L^*)^2}\Delta v_i' + \frac{p^*}{L^*}\frac{\partial p'}{\partial x_i'} = \rho f_i,$$

or, after multiplication by $L^*/(\rho(U^*)^2)$

$$\frac{L*}{t^*U^*}\frac{\partial v_i'}{\partial t'} + 1\cdot\sum_j v_j'\frac{\partial v_i'}{x_j'} - \frac{\mu}{\rho L^*U^*}\Delta v_i' + \frac{p^*}{\rho(U^*)^2}\frac{\partial p'}{\partial x_i'} = f_i'$$

with $f' = L^*/(U^*)^2\, f$. With the setting $t^* = L^*/U^*$, $p^* = \rho(U^*)^2$ and

$$Re = \frac{\rho L^* U^*}{\mu} = \frac{L^* U^*}{\nu},$$

the so-called Reynolds number, one obtains

$$\frac{\partial v}{\partial t} + (v \cdot \operatorname{grad})v - \frac{1}{Re}\Delta v + \operatorname{grad} p = f,$$

and, for the steady-state case:

$$(v \cdot \operatorname{grad})v - \frac{1}{Re}\Delta v + \operatorname{grad} p = f.$$

For $Re \ll 1$ the viscosity of the flow dominates, for $Re \gg 1$ the flow is dominantly convective. For $Re \to \infty$ one formally obtains the so-called Euler equations:

$$\frac{\partial v}{\partial t} + (v \cdot \operatorname{grad})v + \operatorname{grad} p = f.$$

If the transformed equations are multiplied by $(L^*)^2/(\mu U^*)$, one obtains

$$\frac{\rho(L^*)^2}{\mu t^*}\frac{\partial v_i'}{\partial t'} + \frac{\rho L^* U^*}{\mu}\sum_{j=1}^{N} v_j'\frac{\partial v_i'}{x_j'} - 1\cdot\Delta v_i' + \frac{p^* L^*}{\mu U^*}\frac{\partial p'}{\partial x_i'} = f'$$

with $f' = \rho(L^*)^2 f/(\mu U^*)$. With the setting $t^* = (\rho(U^*)^2)/\mu$, $p^* = (\mu U^*)/L^*$ it follows that

$$\frac{\partial v}{\partial t} + Re\,(v\cdot\operatorname{grad})v - \Delta v + \operatorname{grad} p = f,$$

and, for the steady-state case:

$$Re\,(v\cdot\operatorname{grad})v - \Delta v + \operatorname{grad} p = f,$$

In this formulation one obtains for $Re = 0$ the so-called Stokes equations:

$$\frac{\partial v}{\partial t} - \Delta v + \operatorname{grad} p = f.$$

and, for the steady-state case:

$$-\Delta v + \operatorname{grad} p = f.$$

Chapter 2

Variational Problems

2.1 Pure Displacement Problem in Linear(ized) Elasticity

Let $v = (v_1, v_2, v_3)^T$ be a test function from some suitable space V with $v = 0$ on Γ_D. The equilibrium conditions are multiplied component-wise by this test function, are integrated over Ω and are added. Then we obtain:

$$-\int_\Omega \operatorname{div}\sigma \cdot v \; dx = \int_\Omega f \cdot v \; dx.$$

Using integration by parts, we have

$$\int_\Omega \operatorname{div}\sigma \cdot v \; dx = \int_\Gamma \sigma n \cdot v \; ds - \int_\Omega \sigma : \operatorname{grad} v \; dx$$

with the notation

$$A : B = \sum_{i,j=1}^{d} a_{ij} b_{ij}$$

for matrices A, B. Therefore, we obtain

$$\int_\Omega \sigma : \operatorname{grad} v \; dx = \int_\Omega f \cdot v \; dx + \int_\Gamma \sigma n \cdot v \; d\sigma.$$

Since σ is symmetric, we have:

$$\sigma : \operatorname{grad} v = \frac{1}{2}\,\sigma : \operatorname{grad} v + \frac{1}{2}\,\sigma^T : \operatorname{grad} v = \frac{1}{2}\,\sigma : \operatorname{grad} v + \frac{1}{2}\,\sigma : \operatorname{grad} v^T = \sigma : \varepsilon(v).$$

Moreover, since $v = 0$ on Γ_D and $\sigma n = t_N$ on Γ_N:

$$\int_\Gamma \sigma n \cdot v \; ds = \int_{\Gamma_N} \sigma n \cdot v \; ds = \int_{\Gamma_N} t_N \cdot v \; ds.$$

Therefore

$$\int_\Omega \sigma : \varepsilon(v)\ dx = \int_\Omega f \cdot v\ dx + \int_{\Gamma_N} t_N \cdot v\ ds.$$

With Hooke's law

$$\sigma = C\varepsilon(u)$$

we finally obtain the following variational problem:

Find $u \in V_g = \{v \in V : v = u_D \text{ on } \Gamma_D\}$ such that

$$a(u, v) = \langle F, v\rangle \tag{2.1}$$

for all $v \in V_0 = \{v \in V : v = 0 \text{ on } \Gamma_D\}$ with

$$a(u, v) = \int_\Omega C\varepsilon(u) : \varepsilon(v)\ dx$$

and

$$\langle F, v\rangle = \int_\Omega f \cdot v\ dx + \int_{\Gamma_N} t_N \cdot v\ ds.$$

A natural choice for the space V is the Sobolev space $H^1(\Omega, \mathbb{R}^3)$. Observe that $H^1(\Omega, \mathbb{R}^3)$ is a Hilbert space. $\|v\|_1$ denotes the norm in $H^1(\Omega, \mathbb{R}^3)$, $|v|_1$ the semi-norm, built from the first derivatives, and $\|v\|_0$ the L^2-norm:

$$\|v\|_0^2 = \sum_{i=1}^3 \int_\Omega v_i^2\ dx = \int_\Omega v \cdot v\ dx, \quad |v|_1^2 = \sum_{i,j=1}^3 \int_\Omega \left[\frac{\partial v_i}{\partial x_j}\right]^2\ dx = \int_\Omega \operatorname{grad} v : \operatorname{grad} v\ dx$$

and

$$\|v\|_1^2 = \|v\|_0^2 + |v|_1^2.$$

Furthermore, we assume that there is a function $g \in V$ such that $g = u_D$ on Γ_D. Then the problem can be homogenized. Therefore, in the following, we consider, without loss of generality, only homogeneous Dirichlet boundary conditions $u_N = 0$ and set $V = V_g = V_0 = H^1_{0,D}(\Omega, \mathbb{R}^3)$ with

$$H^1_{0,D}(\Omega, \mathbb{R}^3) = \{v \in H^1(\Omega, \mathbb{R}^3) : v = 0 \text{ on } \Gamma_D\}.$$

If $\Gamma_D = \Gamma$ we use the shorter notation $H^1_0(\Omega, \mathbb{R}^3)$ for this space.

For the special case of St.Venant-Kirchhoff materials we obtain:

$$\begin{aligned} a(u, v) &= \int_\Omega [\lambda\ \operatorname{trace}\varepsilon(u)\, I + 2\,\mu\,\varepsilon(u)] : \varepsilon(v)\ dx \\ &= \int_\Omega [\lambda\ \operatorname{trace}\varepsilon(u)\ \operatorname{trace}\varepsilon(v) + 2\mu\,\varepsilon(u) \colon \varepsilon(v)]\ dx \\ &= \int_\Omega [\lambda\ \operatorname{div} u \operatorname{div} v + 2\mu\,\varepsilon(u) \colon \varepsilon(v)]\ dx. \end{aligned}$$

Therefore, the bilinear form a is symmetric

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

and non-negative

$$a(v,v) \geq 0 \quad \text{for all } v \in V.$$

From the symmetry and the non-negativity of the bilinear form a it easily follows that the variational problem (2.1) is equivalent to the following optimization problem:
Find $u \in V_g$ such that

$$J(u) = \inf_{v \in V_g} J(v) \quad \text{with} \quad J(v) = \frac{1}{2}\, a(v,v) - \langle F, v \rangle.$$

Observe that

$$\begin{aligned} \frac{1}{2}\, a(v,v) &= \int_\Omega \left[\frac{\lambda}{2}\, \left(\operatorname{trace} \varepsilon(v)\right)^2 + \mu\, \varepsilon(v) : \varepsilon(v) \right]\, dx \\ &= \int_\Omega \left[\frac{\lambda}{2}\, \left(\operatorname{trace} \varepsilon(v)\right)^2 + \mu\, \operatorname{trace}\left(\varepsilon(v)^2\right) \right]\, dx = \int_\Omega \Psi(\varepsilon(v))\, dx \end{aligned}$$

with the stored energy function Ψ of the St.Venant-Kirchhoff material.

If σ and ε are interpreted as 9-dimensional vectors

$$\begin{aligned} \sigma &= (\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{21}, \sigma_{23}, \sigma_{32}, \sigma_{31}, \sigma_{13})^T, \\ \varepsilon &= (\varepsilon_{11}, \varepsilon_{22}, \varepsilon_{33}, \varepsilon_{12}, \varepsilon_{21}, \varepsilon_{23}, \varepsilon_{32}, \varepsilon_{31}, \varepsilon_{13})^T, \end{aligned}$$

then C in Hooke's law

$$\sigma = C\, \varepsilon,$$

becomes a 9-by-9 matrix. With this interpretation we can write

$$a(u,v) = \int_\Omega C\, \varepsilon(u) \cdot \varepsilon(v)\, dx.$$

In particular, for a St.Venant-Kirchhoff material, we obtain

$$C = \begin{pmatrix} \lambda+2\mu & \lambda & \lambda & & & & & & \\ \lambda & \lambda+2\mu & \lambda & & & & & & \\ \lambda & \lambda & \lambda+2\mu & & & & & & \\ & & & 2\mu & & & & & \\ & & & & 2\mu & & & & \\ & & & & & 2\mu & & & \\ & & & & & & 2\mu & & \\ & & & & & & & 2\mu & \\ & & & & & & & & 2\mu \end{pmatrix}.$$

C has exactly 2 different eigenvalues:

$$\lambda_{\min}(C) = 2\,\mu, \quad \lambda_{\max}(C) = 3\,\lambda + 2\,\mu.$$

So, C is symmetric and positive definite.

The theorem of **Lax-Milgram** is of central importance for showing existence and uniqueness of a solution to (2.1):

Theorem 2.1 (Lax-Milgram)**.** *Let V be a real Hilbert space and assume that*

1. $F \in V^*$.
2. $a\colon V \times V \longrightarrow \mathbb{R}$ *is a bilinear form, which is*
   (a) *bounded on V, i.e.: there is a constant $\mu_2 > 0$ with*
   $$|a(u,v)| \leq \mu_2 \, \|u\|_V \, \|v\|_V \quad \textit{for all } u, v \in V,$$
   and
   (b) *coercive on V, i.e.: there is a constant $\mu_1 > 0$ with*
   $$|a(v,v)| \geq \mu_1 \, \|v\|_V^2 \quad \textit{for all } v \in V.$$

Then the variational problem: find $u \in V$ such that

$$a(u,v) = \langle F, v \rangle \quad \textit{for all } v \in V,$$

has a unique solution and we have

$$\|u\|_V \leq \frac{1}{\mu_1} \|F\|_{V^*}.$$

We will now discuss the assumptions of the theorem of Lax-Migram for the variational problem (2.1) for a St.Venant-Kirchhoff material with

$$f \in L^2(\Omega, \mathbb{R}^3), \quad t_N \in L^2(\Gamma_N, \mathbb{R}^3) \quad \text{and} \quad u_D = 0.$$

1. F is linear: trivial. From the Cauchy inequality it follows:
   $$|\langle F, v \rangle| \leq \|f\|_{0,\Omega} \, \|v\|_{0,\Omega} + \|t_N\|_{0,\Gamma_N} \, \|v\|_{0,\Gamma_N}.$$
   From
   $$\|v\|_{0,\Omega} \leq \|v\|_{1,\Omega} \quad \text{and} \quad \|v\|_{0,\Gamma_N} \leq c(\Gamma_N) \, \|v\|_{1,\Omega}$$
   the boundedness of F follows.
2. a is bilinear: trivial.

(a) a is bounded: We have

$$|C\varepsilon(u) : \varepsilon(v)| \leq \lambda_{\max}(C)\, \|\varepsilon(u)\|_F\, \|\varepsilon(v)\|_F$$

with the Frobenius norm $\|A\|_F = (A : A)^{1/2}$. Therefore,

$$\begin{aligned} |a(u,v)| &\leq \lambda_{\max}(C) \int_\Omega \|\varepsilon(v)\|_F\, \|\varepsilon(u)\|_F\, dx \\ &\leq \lambda_{\max}(C) \left(\int_\Omega \|\varepsilon(u)\|_F^2\, dx \right)^{1/2} \left(\int_\Omega \|\varepsilon(v)\|_F^2\, dx \right)^{1/2}. \end{aligned}$$

Since

$$\left[\frac{1}{2}(a+b) \right]^2 \leq \frac{1}{2}\left(a^2 + b^2\right),$$

it follows that

$$\begin{aligned} \int_\Omega \|\varepsilon(v)\|_F^2\, dx &= \sum_{i,j=1}^3 \int_\Omega \varepsilon_{ij}(v)^2\, dx = \sum_{i,j=1}^3 \int_\Omega \left[\frac{1}{2} \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right) \right]^2 dx \\ &\leq \frac{1}{2} \sum_{i,j=1}^3 \int_\Omega \left[\left(\frac{\partial v_i}{\partial x_j} \right)^2 + \left(\frac{\partial v_j}{\partial x_i} \right)^2 \right] dx = |v|_1^2 \leq \|v\|_1^2. \end{aligned}$$

This implies:

$$|a(u,v)| \leq \lambda_{\max}(C)\, |u|_1\, |v|_1 \leq \lambda_{\max}(C)\, \|u\|_1\, \|v\|_1.$$

(b) a is coercive: We have

$$C\, \varepsilon(v) : \varepsilon(v) \geq \lambda_{\min}(C)\, \varepsilon(v) : \varepsilon(v).$$

Hence

$$a(v,v) = \int_\Omega C\, \varepsilon(v) : \varepsilon(v)\, dx \geq \lambda_{\min}(C) \int_\Omega \varepsilon(v) : \varepsilon(v)\, dx.$$

In order to continue, we need Korn's inequality.

For the case $\Gamma_D = \Gamma$ (first boundary value problem) we need the so-called first Korn inequality:

Lemma 2.1 (First Korn Inequality)**.** *Let $\Omega \subset \mathbb{R}^3$ be open. Then*

$$\int_\Omega \varepsilon(v) \colon \varepsilon(v)\, dx \geq \frac{1}{2}\, |v|_1^2 \quad \textit{for all } v \in H_0^1(\Omega, \mathbb{R}^3).$$

Proof. The set $C_0^\infty(\Omega)$ is dense in $H_0^1(\Omega)$. Therefore, it suffices to show the inequality for all $v \in C_0^\infty(\Omega, \mathbb{R}^3)$:

$$\begin{aligned}\sum_{i,j=1}^{3} \int_\Omega \varepsilon_{ij}(v)\, \varepsilon_{ij}(v)\; dx &= \frac{1}{2} \sum_{i,j=1}^{3} \int_\Omega \varepsilon_{ij}(v) \left(\frac{\partial v_i}{\partial x_j} + \frac{\partial v_j}{\partial x_i} \right)\; dx = \sum_{i,j=1}^{3} \int_\Omega \varepsilon_{ij}(v)\, \frac{\partial v_i}{\partial x_j}\; dx \\ &= \frac{1}{2} \sum_{i,j=1}^{3} \int_\Omega \left(\frac{\partial v_i}{\partial x_j} \right)^2 dx + \frac{1}{2} \sum_{i,j=1}^{3} \int_\Omega \frac{\partial v_i}{\partial x_j}\, \frac{\partial v_j}{\partial x_i}\; dx \\ &= \frac{1}{2} |v|_1^2 + \frac{1}{2} \sum_{i,j=1}^{3} \int_\Omega \frac{\partial v_i}{\partial x_j}\, \frac{\partial v_j}{\partial x_i}\; dx.\end{aligned}$$

Using integration by parts twice it follows:

$$\sum_{i,j=1}^{3} \int_\Omega \frac{\partial v_i}{\partial x_j}\, \frac{\partial v_j}{\partial x_i}\; dx = \sum_{i,j=1}^{3} \int_\Omega \frac{\partial v_i}{\partial x_i}\, \frac{\partial v_j}{\partial x_j}\; dx = \int_\Omega (\operatorname{div} v)^2\; dx \geq 0.$$

This completes the proof. □

This easily implies the coercivity for the first boundary value problem:

$$a(v,v) \geq \lambda_{\min}(C) \int_\Omega \varepsilon(v) \colon \varepsilon(v)\; dx \geq \frac{\lambda_{\min}(C)}{2} |v|_1^2 \geq \frac{\lambda_{\min}(C)}{2(1+c_F^2)} \|v\|_1^2,$$

where c_F denotes the constant from Friedrichs' inequality:

$$\|v\|_0 \leq c_F\, |v|_1,$$

from which it immediately follows that:

$$\|v\|_1^2 \leq (1 + c_F^2) |v|_1^2.$$

For proving the coercivity of the second boundary value problem ($\Gamma_N = \Gamma$) and the mixed boundary value problem ($\Gamma_D \neq \emptyset$ and $\Gamma_N \neq \emptyset$) the second Korn inequality is needed:

Lemma 2.2 (Second Korn Inequality)**.** *Let $\Omega \subset \mathbb{R}^d$ be open and bounded with a Lipschitz-continuous boundary. Then there is a constant $c_K = c_K(\Omega) > 0$ such that*

$$\int_\Omega \varepsilon(v) \colon \varepsilon(v)\; dx + \|v\|_0^2 \geq c_K^2\, \|v\|_1^2 \quad \textit{for all } v \in H^1(\Omega, \mathbb{R}^d).$$

The proof of the second Korn inequality is similar to the proof of the so-called inf-sup condition of the divergence operator, discussed later. For $d = 2$ the statements are even equivalent.

A proof of the second Korn inequality can be found, e.g., in [4], [10].

In order to conclude coercivity from the second Korn inequality, we first need the kernel of $\varepsilon(v)$:

Lemma 2.3. *Let $\Omega \subset \mathbb{R}^3$ be open and connected. Then:*

$$\varepsilon(v) = 0 \Longleftrightarrow v(x) = a \times x + b$$

with some constant vectors $a, b \in \mathbb{R}^3$.

Proof. Assume $\varepsilon(v) = 0$. Then we have (in $H^{-1}(\Omega)$):

$$\frac{\partial^2}{\partial x_i \partial x_j} v_k = \frac{\partial}{\partial x_i} \varepsilon_{jk}(v) + \frac{\partial}{\partial x_j} \varepsilon_{ik}(v) - \frac{\partial}{\partial x_k} \varepsilon_{ij}(v) = 0.$$

Therefore, v is a linear function:

$$v(x) = Ax + b,$$

where A is a real 3-by-3 matrix. Hence

$$\varepsilon(v) = \frac{1}{2}\left(A + A^T\right)$$

and

$$\varepsilon(v) = 0 \Longleftrightarrow A = -A^T \Longleftrightarrow A = \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}.$$

Since

$$Ax = a \times x$$

with $a = (a_1, a_2, a_3)^T$, the proof is completed. The reverse direction is trivial. □

Now the coercivity can be shown for the mixed and the second boundary value problem:

Corollary 2.1. *Let $\Omega \subset \mathbb{R}^3$ be an open, bounded and connected domain with Lipschitz-continuous boundary $\Gamma = \partial\Omega$. Then:*

1. *If $\Gamma_D \subset \Gamma$ with $\operatorname{meas}_2(\Gamma_D) \neq 0$, then there exists a constant $c_K = c_K(\Omega) > 0$ such that*

$$\int_\Omega \varepsilon(v) \colon \varepsilon(v)\ dx \geq c_K^2\, |v|_1^2$$

   for all $v \in V = H^1_{0,D}(\Omega, \mathbb{R}^3)$.

2. *If $\Gamma_D = \emptyset$, then there exists a constant $c_K = c_K(\Omega) > 0$ with*

$$\int_\Omega \varepsilon(v) \colon \varepsilon(v)\ dx \geq c_K^2\, |v|_1^2$$

   for all $v \in V = \hat{H}(\Omega) = \{v \in V = H^1(\Omega, \mathbb{R}^3) \mid \int_\Omega v\ dx = 0,\ \int_\Omega \operatorname{curl} v\ dx = 0\}$.

Proof. Assume that the inequality does not hold. Then there is a sequence (v_n) in V with

$$\int_\Omega \varepsilon(v_n) \colon \varepsilon(v_n)\, dx \to 0 \quad \text{and} \quad |v_n|_1 = 1.$$

From Friedrichs' inequality or Poincaré's inequality it follows that there is a constant $c > 0$ with

$$\|v_n\|_1 \leq c\, |v_n|_1 = c \quad \text{for all } n \in \mathbb{N}.$$

Hence (v_n) is a bounded sequence in $H^1(\Omega, \mathbb{R}^3)$.

The embedding $H^1(\Omega, \mathbb{R}^3) \longrightarrow L^2(\Omega, \mathbb{R}^3)$ is compact. Therefore, there exists a subsequence $(v_{n'})$ which converges in $L^2(\Omega, \mathbb{R}^3)$.

The second Korn inequality implies

$$\begin{aligned} c_K^2\, \|v_{n'} - v_{m'}\|_1^2 &\leq \int_\Omega \varepsilon(v_{n'} - v_{m'}) \colon \varepsilon(v_{n'} - v_{m'})\, dx + \|v_{n'} - v_{m'}\|_0^2 \\ &\leq 2 \int_\Omega \varepsilon(v_{n'}) \colon \varepsilon(v_{n'})\, dx + 2 \int_\Omega \varepsilon(v_{m'}) : \varepsilon(v_{m'})\, dx + \|v_{n'} - v_{m'}\|_0^2 \longrightarrow 0 \end{aligned}$$

for $n', m' \to \infty$.

So $(v_{n'})$ converges in $H^1(\Omega, \mathbb{R}^3)$ towards some element v_0. Then, however:

$$\varepsilon(v_0) = \lim_{n' \to \infty} \varepsilon(v_{n'}) = 0$$

and, therefore, $v_0 = 0$ because of the definition of V and Lemma 2.3 in contradiction to

$$|v_0|_1 = \lim_{n' \to \infty} |v_{n'}|_1 = 1.$$

□

In summary, we have

Corollary 2.2. *Under the assumptions of Lemma 2.1 and Corollary 2.1 the bilinear form a is coercive on V with*

$$a(v, v) \geq \mu_1\, |v|_1^2$$

where

$$\mu_1 = \lambda_{min}(C)\, c_K^2.$$

So all assumptions of the theorem of Lax-Milgram are satisfied and we have:

Theorem 2.2. *Under the appropriate assumptions the formulated boundary value problems in linear(ized) elasticity are well-posed.*

Remark: In the case of pure Neumann boundary conditions the so-called compatibility conditions

$$\int_\Omega f\ dx + \int_\Gamma t_N\ ds = 0 \quad \text{and} \quad \int_\Omega x \times f\ dx + \int_\Gamma x \times t_N\ ds = 0,$$

are necessary and sufficient that a solution of the variational problem in V is also a solution of the variational problem in $H^1(\Omega, \mathbb{R}^3)$. The solution in V is unique up to an arbitrary element from the kernel of $\varepsilon(v)$.

For estimating the discretization error or the condition number of the stiffness matrix for finite element methods the ratio μ_2/μ_1 (the condition number of the problem) is of essential importance. Using $|v|_1$ as the norm in V, which is equivalent to $\|v\|_1$ by Friedrichs' or Poincaré's inequality, we obtain the following estimate for this condition number

$$\frac{\mu_2}{\mu_1} \leq \frac{\lambda_{\max}(C)}{\lambda_{\min}(C)} \frac{1}{c_K^2} = \kappa(C) \frac{1}{c_K^2},$$

where $\kappa(C) = \lambda_{\max}(C)/\lambda_{\min}(C)$ denotes the condition number of C .

For certain values of the data (Ω, Γ_D, Γ_N, C, f and g) the condition number μ_2/μ_1 can become very large, e.g.:

1. Almost incompressible materials: For $\nu \to 1/2$ we have:

$$\kappa(C) = \frac{3\,\lambda + 2\,\mu}{2\,\mu} = \frac{1+\nu}{1-2\,\nu} \to \infty.$$

   This is called material locking.

2. Long cantilever (Kragbalken):

   In this case the constant in Korn's inequality is very small:

$$c_K^{-1} = \sup_{v \in V} \frac{|v|_1}{\|\varepsilon(v)\|_0} \geq \sqrt{1 + 2\left(\frac{L}{H}\right)^2} = O\left(\frac{L}{H}\right) \longrightarrow \infty$$

   for $H \ll L$. (Choose $v(x, y, z) = (2xy, -x^2, 0)^T \in V$.) This phenomenon is called geometry locking.

In the next section we will reformulate the linear elasticity problem as a mixed variational problem. In this setting material locking for $\nu \to 1/2$ can be avoided.

2.2 Mixed Variational Problems in Continuum Mechanics

2.2.1 Incompressible and Almost Incompressible Materials

We consider only the case of a St.Venant-Kirchhoff material. The (primal) variational problem reads:

Find $u \in V_g$, such that

$$\int_\Omega [\lambda \operatorname{div} u \operatorname{div} v + 2\,\mu\,\varepsilon(u)\colon \varepsilon(v)]\ dx = \int_\Omega f \cdot v\ dx + \int_{\Gamma_N} t_N \cdot v\ ds$$

for all $v \in V_0$.

For $\lambda \longrightarrow \infty$, i.e.: $\nu = \lambda/(2(\lambda+\mu)) \longrightarrow 1/2$ the problem becomes very ill-conditioned. The basic idea is to derive a so-called mixed variational formulation by introducing a new variable

$$p = \lambda \operatorname{div} u.$$

Then

$$2\mu \int_\Omega \varepsilon(u) : \varepsilon(v)\ dx + \int_\Omega p \operatorname{div} v\ dx = \int_\Omega f \cdot v\ dx + \int_{\Gamma_N} t_N \cdot v\ ds$$

for all $v \in V_0$ and

$$\int_\Omega q \operatorname{div} u\ dx - \frac{1}{\lambda} \int_\Omega p\,q\ dx = 0$$

for all $q \in L^2(\Omega)$. So the following mixed variational problem results:

Find $u \in V_g$ and $p \in L^2(\Omega)$, such that

$$\begin{array}{rcll} a(u,v) + b(v,p) & = & \langle F, v\rangle & \text{for all } v \in V_0, \\ b(u,q) - t^2\, c(p,q) & = & 0 & \text{for all } q \in L^2(\Omega) \end{array}$$

with

$$a(u,v) = 2\mu \int_\Omega \varepsilon(u) : \varepsilon(v)\ dx, \quad b(v,p) = \int_\Omega p \operatorname{div} v\ dx, \quad c(p,q) = \int_\Omega p\,q\ dx$$

and

$$\langle F, v\rangle = \int_\Omega f \cdot v\ dx + \int_{\Gamma_N} t_N \cdot v\ ds, \quad t^2 = \frac{1}{\lambda}.$$

For the limit case $t = 0$ the following variational problem is obtained:

Find $u \in V_g$ and $p \in L^2(\Omega)$, such that

$$\begin{array}{rcll} a(u,v) + b(v,p) & = & \langle F, v\rangle & \text{for all } v \in V_0, \\ b(u,q) & = & 0 & \text{for all } q \in L^2(\Omega) \end{array}$$

for describing incompressible materials.

2.2.2 The Stokes Problem in Fluid Mechanics

Consider the steady state Stokes problem in some domain Ω:

$$\begin{aligned} -\nu\,\Delta u + \operatorname{grad} p &= f \quad \text{in } \Omega, \\ \operatorname{div} u &= 0 \quad \text{in } \Omega, \end{aligned}$$

where, instead of the original notation v from now on the velocity is denoted by u and, for simplicity we set $\rho = 1$.

Here we will consider only the boundary condition:

$$u = u_D, \quad x \in \Gamma.$$

Let v be a test function with $v = 0$ on Γ. By multiplying the balance law of momentum by v and integrating over Ω we obtain:

$$-\nu \int_\Omega \Delta u \cdot v \; dx + \int_\Omega \operatorname{grad} p \cdot v \; dx = \int_\Omega f \cdot v \; dx.$$

By integration by parts it follows:

$$\int_\Omega \Delta u \cdot v \; dx = \int_{\partial\Omega} \frac{\partial u}{\partial n} \cdot v \; ds - \int_\Omega \operatorname{grad} u : \operatorname{grad} v \; dx = -\int_\Omega \operatorname{grad} u : \operatorname{grad} v \; dx$$

and

$$\int_\Omega \operatorname{grad} p \cdot v \; dx = \int_{\partial\Omega} p \; v \cdot n \; ds - \int_\Omega p \; \operatorname{div} v \; dx = -\int_\Omega p \; \operatorname{div} v \; dx.$$

Therefore, we obtain the following weak form of the balance law of momentum:

$$\nu \int_\Omega \operatorname{grad} u : \operatorname{grad} v \; dx - \int_\Omega p \; \operatorname{div} v \; dx = \int_\Omega f \cdot v \; dx.$$

The weak form of the law of continuity is obtained by multiplying with an arbitrary test function $q \in L^2(\Omega)$ and integrating over Ω:

$$\int_\Omega q \; \operatorname{div} u \; dx = 0.$$

In summary the weak or variational form of the Stokes equation reads:

Find $u \in V_g$ and $p \in L^2(\Omega)$, such that

$$\begin{aligned} a(u,v) + b(v,p) &= \langle F, v \rangle \quad \text{for all } v \in V_0, \\ b(u,q) \qquad\quad &= 0 \qquad\quad \text{for all } q \in L^2(\Omega) \end{aligned}$$

with

$$a(u,v) = \nu \int_\Omega \operatorname{grad} u : \operatorname{grad} v \; dx, \quad b(v,q) = -\int_\Omega q \; \operatorname{div} v \; dx,$$

$$\langle F, v \rangle = F(v) = \int_\Omega f \cdot v \; dx$$

and the spaces

$$V = H^1(\Omega, \mathbb{R}^3), \quad V_0 = H_0^1(\Omega, \mathbb{R}^3), \quad V_g = \{v \in V : v = u_D \text{ on } \Gamma\}.$$

2.2.3 The Hellinger-Reissner Formulation for Linear Elasticity

Starting point is the following classical formulation: Find the displacement u and the stress σ, such that:

$$\begin{aligned} C^{-1}\,\sigma - \varepsilon(u) &= 0 \quad && \text{in } \Omega, \\ \operatorname{div}\sigma &= -f && \text{in } \Omega, \\ u &= u_D && \text{on } \Gamma_D, \\ \sigma\, n &= t_N && \text{on } \Gamma_N. \end{aligned}$$

Let τ be a mapping from Ω to $\mathbb{S} = \mathbb{R}^{3\times 3}_{\text{sym}}$, the space of symmetric 3-by-3 matrices. By multiplying the first equation component-wise by the test function τ, integrating over Ω and adding, we obtain:

$$\int_\Omega C^{-1}\sigma : \tau \; dx - \int_\Omega \tau : \varepsilon(u) \; dx = 0.$$

Let v be a test function mapping from $\overline{\Omega}$ to $\mathbb{R}^3$ with $v = 0$ on Γ_D. By multiplying the second equation component-wise by v, integrating over Ω, and adding, we obtain (after integration by parts):

$$-\int_\Omega \sigma : \varepsilon(v) \; dx = -\int_\Omega f \cdot v \; dx - \int_{\Gamma_N} t_N \cdot v \; ds.$$

Therefore, the following mixed variational problem results:

Find $\sigma \in L^2(\Omega, \mathbb{S})$ and $u \in V_g \subset H^1(\Omega, \mathbb{R}^3)$ (see the primal variational problem) such that

$$\begin{aligned} a(\sigma,\tau) + b(\tau,u) &= 0 && \text{for all } \tau \in L^2(\Omega,\mathbb{S}), \\ b(\sigma, v) &= \langle G, v\rangle && \text{for all } v \in V_0 \end{aligned}$$

with

$$a(\sigma,\tau) = \int_\Omega C^{-1}\sigma : \tau \; dx, \quad b(\tau,u) = -\int_\Omega \tau : \varepsilon(u) \; dx$$

and

$$\langle G, v\rangle = -\int_\Omega f \cdot v \; dx - \int_{\Gamma_N} t_N \cdot v \; dx.$$

The norm $\|.\|_{L^2(\Omega,\mathbb{S})}$ (or, in short $\|.\|_0$) in the space $L^2(\Omega,\mathbb{S})$ is given by

$$\|\tau\|^2_{L^2(\Omega,\mathbb{S})} = \sum_{i,j=1}^{3} \|\tau_{ij}\|_0^2.$$

Another variational formulation is obtained by using integration by parts for the second term in the first equation:

$$\int_\Omega \tau : \varepsilon(u) \; dx = \int_\Omega \tau : \operatorname{grad} u \; dx = \int_\Gamma \tau\, n \cdot u \; ds - \int_\Omega \operatorname{div}\tau \cdot u \; dx$$

Then we obtain for test functions τ mapping from $\overline{\Omega}$ to $\mathbb{S}$ with $\tau\, n = 0$ on Γ_N:

$$\int_\Omega C^{-1}\sigma : \tau \; dx + \int_\Omega \operatorname{div} \tau \cdot u \; dx = \int_{\Gamma_D} \tau\, n \cdot u_D \; ds.$$

Without using integration by parts the second equation reads for arbitrary test functions v mapping Ω to $\mathbb{R}^3$:

$$\int_\Omega \operatorname{div} \sigma \cdot v \; dx = -\int_\Omega f \cdot v \; dx.$$

Then the following mixed variational problem results:

Find $\sigma \in V_g$ and $u \in Q = L^2(\Omega, \mathbb{R}^3)$, such that

$$\begin{aligned} a(\sigma,\tau) + b(\tau,u) &= \langle F, \tau\rangle \quad \text{for all } \tau \in V_0, \\ b(\sigma, v) &= \langle G, v\rangle \quad \text{for all } v \in Q \end{aligned}$$

with

$$a(\sigma,\tau) = \int_\Omega C^{-1}\sigma : \tau \; dx, \quad b(\tau,u) = \int_\Omega \operatorname{div}\tau \cdot u \; dx$$

and

$$\langle F, \tau\rangle = \int_{\Gamma_D} \tau\, n \cdot u_D \; ds, \quad \langle G, v\rangle = -\int_\Omega f \cdot v \; dx$$

and the spaces

$$V = \{\tau = (\tau_{ij}) \in L^2(\Omega, \mathbb{S}) : \operatorname{div}\tau \in L^2(\Omega, \mathbb{R}^3)\} = H(\operatorname{div}, \Omega, \mathbb{S}),$$

$$\begin{aligned} V_0 &= \{\tau \in V \; : \; "\tau\, n = 0 \text{ on } \Gamma_N"\} \\ &= \{\tau \in V \; : \; \langle \tau\, n, v\rangle = 0 \text{ for all } v \in H^1_{0,D}(\Omega, \mathbb{R}^3)\} = H_{0,N}(\operatorname{div}, \Omega, \mathbb{S}) \end{aligned}$$

and

$$\begin{aligned} V_g &= \{\tau \in V \; : \; "\tau\, n = t_N \text{ on } \Gamma_N"\} \\ &= \{\tau \in V \; : \; \langle \tau\, n, v\rangle = \langle t_N, v\rangle \text{ for all } v \in H^1_{0,D}(\Omega, \mathbb{R}^3)\}. \end{aligned}$$

The norm $\|.\|_{H(\operatorname{div},\Omega,\mathbb{S})}$ in the space V is given by

$$\|\tau\|^2_{H(\operatorname{div},\Omega,\mathbb{S})} = \|\tau\|^2_0 + \|\operatorname{div}\tau\|^2_0.$$

It can be shown that V is a Hilbert space and the trace $\tau\, n$ is well-defined in V.

Remark:

1. Observe that for the second variant of the Hellinger-Reissner formulation the boundary condition

$$u = u_D \quad \text{on } \Gamma_D$$

is a natural boundary condition, while

$$\sigma\, n = t_N \quad \text{on } \Gamma_N$$

is an essential boundary condition. For the original primal variational formulation and the first form of the mixed variational formulation the situation is the opposite.

2. In applications it is often more important to obtain accurate information on stresses than on the displacement. The second variant of the Hellinger-Reissner formulation helps in this direction.

3. For the case of pure Dirichlet boundary conditions ($\Gamma_N = \emptyset$) one chooses the space

$$V_0 = \{\tau \in H(\mathrm{div}, \Omega, \mathbb{S}) : \int_\Omega \mathrm{trace}\, \tau \; dx = 0\}$$

for the second variant of the Hellinger-Reissner formulation to ensure uniqueness of the solution.

2.3 The Theorems of Brezzi and Babuška-Aziz

All mixed variational problems considered so far have the following form (after homogenization):

Find $u \in V$ and $p \in Q$ such that

$$\begin{aligned} a(u,v) + b(v,p) &= \langle F, v\rangle \quad \text{for all } v \in V, \\ b(u,q) - t^2\, c(p,q) &= \langle G, q\rangle \quad \text{for all } q \in Q, \end{aligned}$$

where V and Q are suitable Hilbert spaces, $a : V \times V \longrightarrow \mathbb{R}$, $b : V \times Q \longrightarrow \mathbb{R}$ and $c : Q \times Q \longrightarrow \mathbb{R}$ are bounded bilinear forms, $F : V \longrightarrow \mathbb{R}$ and $G : Q \longrightarrow \mathbb{R}$ are bounded linear functionals, and t is a real parameter with $t \geq 0$.

In the special case $t = 0$ we obtain the problem

Find $u \in V$ and $p \in Q$ such that

$$\begin{aligned} a(u,v) + b(v,p) &= \langle F, v\rangle \quad \text{for all } v \in V, \\ b(u,q) &= \langle G, q\rangle \quad \text{for all } q \in Q. \end{aligned}$$

Remark: Let V^* and Q^* be the dual spaces of V and Q. The operators $A : V \longrightarrow V^*$, $B : V \longrightarrow Q^*$, $B^* : Q \longrightarrow V^*$, and $C : Q \longrightarrow Q^*$ are defined by

$$\langle Au, v\rangle = a(u, v), \quad \langle Bv, q\rangle = b(v, q), \quad \langle B^*q, v\rangle = b(v, q), \quad \langle Cp, q\rangle = c(p, q).$$

The operator B^* is called the adjoint (operator) of B.

Then we obtain the following representation of the mixed variational problem as operator equations:

$$\begin{aligned} Au + B^*v &= F, \\ Bu - t^2\, Cp &= G. \end{aligned}$$

Remark: The mixed variational problem can also be formulated as a non-mixed variational problem on $V \times Q$:

Find $(u, p) \in V \times Q$, such that

$$\mathcal{B}((u, p), (v, q)) = \langle F, v\rangle + \langle G, q\rangle \quad \text{for all } (v, q) \in V \times Q$$

with

$$\mathcal{B}((u, p), (v, q)) = a(u, v) + b(v, p) + b(u, q) - t^2\ c(p, q).$$

Observe that $\mathcal{B}$ cannot be coercive for non-negative bilinear forms c:

$$\mathcal{B}((0, q), (0, q)) = -t^2\ c(q, q) \leq 0.$$

Therefore, the theorem of Lax-Milgram is not applicable.

Remark: If, in addition, a and c are symmetric and non-negative bilinear forms, then the mixed variational problem can be formulated as a saddle point problem:

Find $(u, p) \in V \times Q$, such that

$$\mathcal{L}(u, q) \leq \mathcal{L}(u, p) \leq \mathcal{L}(v, p) \quad \text{for all } (v, q) \in V \times Q$$

with

$$\mathcal{L}(v, q) = \frac{1}{2}a(v, v) + b(v, q) - \frac{t^2}{2}c(q, q) - \langle F, v\rangle - \langle G, q\rangle.$$

Remark: If C is invertible, then one obtains the following equivalent unconstrained optimization problem for $t > 0$:

Find $u \in V$, such that

$$J_t(u) = \inf_{v \in V} J_t(v)$$

with

$$J_t(v) = \frac{1}{2}a(v, v) - \langle F, v\rangle + \frac{1}{2t^2}\langle Bv - G, C^{-1}(Bv - G)\rangle.$$

This can be interpreted as a penalty method for solving the constrained optimization problem:

Find $u \in V$, such that

$$J(u) = \inf_{v \in V_g} J(v)$$

with

$$J(v) = \frac{1}{2} a(v, v) - \langle F, v \rangle$$

and

$$V_g = \{v \in V \mid b(v, q) = \langle G, q \rangle \text{ for all } q \in Q\}.$$

The next theorem is of central importance:

Theorem 2.3 (Closed Range Theorem)**.** *Let X and Y be real Hilbert spaces, $\mathcal{A}\colon X \longrightarrow Y^*$ be a linear continuous operator and $\mathcal{A}^*\colon Y \longrightarrow X^*$ be the adjoint operator, given by $\langle \mathcal{A}^* y, x \rangle = \langle \mathcal{A} x, y \rangle$. Then the following statements are equivalent:*

1. $\operatorname{Im} \mathcal{A}$ *is closed;*
2. $\operatorname{Im} \mathcal{A}^*$ *is closed;*
3. $\operatorname{Im} \mathcal{A} = (\operatorname{Ker} \mathcal{A}^*)^\circ$*;*
4. $\operatorname{Im} \mathcal{A}^* = (\operatorname{Ker} \mathcal{A})^\circ$.

The following notations were used: $W^\circ \subset Z^*$ denotes the polar of the sub-space $W \subset Z$:

$$W^\circ = \{l \in Z^* \mid \langle l, w \rangle = 0 \text{ for all } w \in W\}.$$

An immediate consequence of this theorem is:

Corollary 2.3. *Let X and Y be real Hilbert spaces, $\mathcal{A}\colon X \longrightarrow Y^*$ be a linear and continuous operator, $\mathsf{a}\colon X \times Y \longrightarrow \mathbb{R}$ be the corresponding bilinear form, and $\mathcal{A}^*\colon Y \longrightarrow X^*$ be the adjoint operator, given by $\langle \mathcal{A}^* y, x \rangle = \mathsf{a}(x, y) = \langle \mathcal{A} x, y \rangle$. Then the following statements are equivalent:*

1. *There is a constant $\mu_1 > 0$ with*

$$\inf_{0 \neq x \in X} \sup_{0 \neq y \in Y} \frac{\mathsf{a}(x, y)}{\|x\|_X \, \|y\|_Y} \geq \mu_1 > 0. \tag{2.2}$$

2. $\mathcal{A}\colon X \longrightarrow (\operatorname{Ker} \mathcal{A}^*)^\circ$ *is an isomorphism, and there is a constant $\mu_1 > 0$ with*

$$\|\mathcal{A} x\|_{Y^*} \geq \mu_1 \|x\|_X \quad \text{for all } x \in X. \tag{2.3}$$

3. $\mathcal{A}^*\colon (\operatorname{Ker}\mathcal{A}^*)^\perp \longrightarrow X^*$ *is an isomorphism, and there is a constant* $\mu_1 > 0$ *with*

$$\|\mathcal{A}^* y\|_{X^*} \geq \mu_1 \|y\|_Y \quad \textit{for all } y \in (\operatorname{Ker}\mathcal{A}^*)^\perp. \tag{2.4}$$

3.' There is a constant $\mu_1 > 0$ *such that, for each* $x^* \in X^*$*, there exists a* $y \in Y$ *with*

$$\mathcal{A}^* y = x^* \quad \textit{and} \quad \|y\|_Y \leq \frac{1}{\mu_1}\|x^*\|_{X^*}.$$

*3." *$\mathcal{A}^*\colon Y \longrightarrow X^*$ *is surjective.*

Proof.

(1) $\Longrightarrow$ (2): From (2.2) = (2.3) it follows that $\mathcal{A}$ is injective and that $\operatorname{Im}\mathcal{A}$ is closed:

Let $(\mathcal{A}x_n)_{n\in\mathbb{N}}$ be a convergent sequence in $\operatorname{Im}\mathcal{A}$, i.e. $\mathcal{A}x_n \longrightarrow y^*$, then it follows because of (2.2) = (2.3), that the sequence $(x_n)_{n\in\mathbb{N}}$ also converges in Y: $x_n \longrightarrow x$ and $\mathcal{A}x_n \longrightarrow \mathcal{A}x$ since $\mathcal{A}$ is continuous. Hence $y^* = \mathcal{A}x \in \operatorname{Im}\mathcal{A}$.

Then it follows from the Closed Range Theorem that $\operatorname{Im}\mathcal{A} = (\operatorname{Ker}\mathcal{A}^*)^\circ$. So, $\mathcal{A}\colon x \longrightarrow (\operatorname{Ker}\mathcal{A}^*)^\circ$ is bijective and continuous. The inverse mapping is continuous because of (2.2) = (2.3).

(2) $\Longrightarrow$ (1): trivial

(2) $\Longleftrightarrow$ (3): The equivalence of the isomorphism of $\mathcal{A}$ and $\mathcal{A}^*$ directly follows from the Closed Range Theorem. It remains to show the equivalence of the inequalities:

Assume (2.3) and let $y \in (\operatorname{Ker}\mathcal{A}^*)^\perp$. Then $(y,.)_Y \in (\operatorname{Ker}\mathcal{A}^*)^\circ$. Therefore, a $x \in X$ exists with $\mathcal{A}x = (y,.)_Y$ because of (2) and it follows:

$$\|\mathcal{A}^* y\|_{X^*} \geq \frac{\mathsf{a}(x,y)}{\|x\|_X} = \frac{\|y\|_Y^2}{\|x\|_X} = \frac{\|(y,.)_X\|_{X^*}}{\|x\|_X}\|y\|_Y = \frac{\|\mathcal{A}x\|_{Y^*}}{\|x\|_X}\|y\|_Y \geq \mu_1\|y\|_Y.$$

Assume (2.4) and let $x \in X$, then $(x,.)_X \in X^*$. Therefore, a $y \in (\operatorname{Ker}\mathcal{A}^*)^\perp$ exists with $\mathcal{A}^* y = (x,.)_X$ and it follows:

$$\|\mathcal{A}x\|_{Y^*} \geq \frac{\mathsf{a}(x,y)}{\|y\|_Y} = \frac{\|x\|_X^2}{\|y\|_Y} = \frac{\|(x,.)_X\|_{X^*}}{\|y\|_Y}\|x\|_X = \frac{\|\mathcal{A}^*y\|_{X^*}}{\|y\|_Y}\|x\|_X \geq \mu_1\|x\|_X.$$

(3) $\Longrightarrow$ (3′): For y choose the unique solution of $\mathcal{A}^* y = x^*$ in $(\operatorname{Ker}\mathcal{A}^*)^\perp$.

(3′) $\Longrightarrow$ (3): (3') implies that $\mathcal{A}^*$ is surjective. It remains to show that (2.4) is satisfied: Let $y \in (\operatorname{Ker}\mathcal{A}^*)^\perp$ and set $x^* = \mathcal{A}^* y$. From (3') it follows that there exists a $\tilde{y} \in Y$ with $\mathcal{A}^*\tilde{y} = x^*$ and

$$\|\tilde{y}\|_Y \leq \frac{1}{\mu_1}\|x^*\|_{X^*} = \frac{1}{\mu_1}\|\mathcal{A}^* y\|_{X^*}.$$

Since $\tilde{y} - y \in \operatorname{Ker}\mathcal{A}^*$ and $y \in (\operatorname{Ker}\mathcal{A}^*)^\perp$, we have $\|y\|_Y \leq \|\tilde{y}\|_Y$.

(3) $\Longrightarrow$ (3″): trivial

(3″) $\Longrightarrow$ (3): By the Open Mapping Theorem it follows that the inverse of a bijective mapping is continuous, which implies the existence of μ_1. □

Now the theorem of Babuška and Aziz can be easily shown:

Theorem 2.4 (Babuška and Aziz)**.** *Let X and Y be real Hilbert spaces, $\mathcal{A}\colon X \longrightarrow Y^*$ be a linear and continuous operator with corresponding bilinear form $\mathsf{a}\colon X \times Y \longrightarrow \mathbb{R}$, given by $\mathsf{a}(x,y) = \langle \mathcal{A}x, y\rangle$. Then $\mathcal{A}$ is an isomorphism if and only if the following conditions are satisfied:*

1. *There exists a constant $\mu_2 \geq 0$ with*

$$|\mathsf{a}(x,y)| \leq \mu_2 \|x\|_X \|y\|_Y \quad \text{for all } x \in X,\ y \in Y,$$

   i.e.:

$$\|Ax\|_{Y^*} \leq \mu_2 \|x\|_X \quad \text{for all } x \in X.$$

2. *There exists a constant $\mu_1 > 0$ with*

$$\inf_{0\neq x\in X} \sup_{0\neq y\in Y} \frac{\mathsf{a}(x,y)}{\|x\|_X \|y\|_Y} \geq \mu_1,$$

   i.e.:

$$\|Ax\|_{Y^*} \geq \mu_1 \|x\|_X \quad \text{for all } x \in X.$$

3. *For each $y \in Y$ with $y \neq 0$ there exists a $x \in X$ with*

$$\mathsf{a}(x,y) \neq 0,$$

   i.e.:

$$\operatorname{Ker} \mathcal{A}^* = \{0\}.$$

Proof. The statement immediately follows from Corollary 2.3. □

Remark: From the theorem of Babuška-Aziz the theorem of Lax and Milgram follows: Let $X = Y = V$, then the coercivity of a implies the second condition:

$$\sup_{v\in V} \frac{a(u,v)}{\|v\|_V} \geq \frac{a(u,u)}{\|u\|_V} \geq \mu_1 \|u\|_V.$$

The third condition also follows from the coercivity: For $v \neq 0$ choose $u = v$. Then:

$$a(u,v) = a(v,v) \geq \mu_1 \|v\|_V^2 > 0.$$

For mixed variational problems existence and uniqueness of a solution follow from the **theorem of Brezzi**:

Theorem 2.5 (Brezzi)**.** *Let V and Q be real Hilbert spaces, $F \in V^*$, $G \in Q^*$, $a\colon V\times V \longrightarrow \mathbb{R}$ and $b\colon V\times Q \longrightarrow \mathbb{R}$ be bilinear forms. Assume that there exist constants $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$ with*

1. $|a(u,v)| \le \alpha_2 \|u\|_V \|v\|_V$ *for all* $u, v \in V$,
2. $|b(v,q)| \le \beta_2 \|v\|_V \|q\|_Q$ *for all* $v \in V$, $q \in Q$,
3. $a(v,v) \ge \alpha_1 \|v\|_V^2$ *for all* $v \in W = \operatorname{Ker} B = \{v \in V : b(v,q) = 0 \text{ for all } q \in Q\}$,
4. $\inf\limits_{0 \ne q \in Q} \sup\limits_{0 \ne v \in V} \dfrac{b(v,q)}{\|v\|_V \ \|q\|_Q} \ge \beta_1 > 0.$

Then the variational problem
Find $u \in V$ *and* $p \in Q$, *such that*

$$\begin{array}{lcll} a(u,v) + b(v,p) & = & \langle F, v\rangle & \text{for all } v \in V \\ b(u,q) & = & \langle G, q\rangle & \text{for all } q \in Q \end{array}$$

has a unique solution and we have:

$$\begin{aligned} \|u\|_V &\le \frac{1}{\alpha_1}\|F\|_{V^*} + \frac{1}{\beta_1}\left(1 + \frac{\alpha_2}{\alpha_1}\right)\|G\|_{Q^*} \\ \|p\|_Q &\le \frac{1}{\beta_1}\left(1 + \frac{\alpha_2}{\alpha_1}\right)\|F\|_{V^*} + \frac{\alpha_2}{\beta_1^2}\left(1 + \frac{\alpha_2}{\alpha_1}\right)\|G\|_{Q^*} \end{aligned}$$

Proof. Condition (4) corresponds to the condition (1) in Corollary 2.3 for

$$X = Q, \quad Y = V, \quad \mathcal{A} = B^*.$$

Because of Corollary 2.3 (3) there is a unique $g \in W^\perp$ with

$$Bg = G$$

and we have

$$\|g\|_V \le \frac{1}{\beta_1}\|G\|_{Q^*}.$$

Let $w \in W$ be the unique solution of the variational problem

$$a(w,v) = \langle F, v\rangle - a(g,v), \quad \text{for all } v \in W.$$

From the theorem of Lax-Milgram it follows

$$\|w\|_V \le \frac{1}{\alpha_1}\left(\|F\|_{V^*} + \alpha_2\|g\|_V\right).$$

Finally, from Corollary 2.3 (2) it follows that there exists a unique solution $p \in Q$ of the equation

$$B^* p = F - Au$$

with $u = g + w$, since

$$\langle F - Au, w\rangle = 0 \quad \text{for all } w \in W, \quad \text{so} \quad F - Au \in W^\circ.$$

and we obtain

$$\|p\|_Q \leq \frac{1}{\beta_1}\left(\|F\|_{V^*} + \alpha_2\|u\|_V\right).$$

Then $u \in V$ and $p \in Q$ solve the mixed variational problem and the following estimates hold:

$$\begin{aligned}\|u\|_V &\leq \|g\|_V + \|w\|_V \leq \|g\|_V + \frac{1}{\alpha_1}\left(\|F\|_{V^*} + \alpha_2\|g\|_V\right)\\ &\leq \frac{1}{\alpha_1}\|F\|_{V^*} + \frac{1}{\beta_1}\left(1+\frac{\alpha_2}{\alpha_1}\right)\|G\|_{Q^*}\end{aligned}$$

and

$$\begin{aligned}\|p\|_Q &\leq \frac{1}{\beta_1}\left(\|F\|_{V^*} + \alpha_2\|u\|_V\right)\\ &\leq \frac{1}{\beta_1}\left(1+\frac{\alpha_2}{\alpha_1}\right)\|F\|_{V^*} + \frac{\alpha_2}{\beta_1^2}\left(1+\frac{\alpha_2}{\alpha_1}\right)\|G\|_{Q^*}.\end{aligned}$$

□

Remark: If, in addition, a is a symmetric bilinear form, then it can be shown that the mixed variational problem

Find $u \in V$, such that

$$\begin{aligned}a(u,v) + b(v,p) &= \langle F, v\rangle \qquad \text{for all } v \in V\\ b(u,q) &= \langle G, q\rangle \qquad \text{for all } q \in Q\end{aligned}$$

is equivalent to the optimization problem:

Find $u \in W_g$, such that

$$J(u) = \inf_{v \in W_g} J(v)$$

with

$$J(v) = \frac{1}{2}a(v,v) - \langle F, v\rangle$$

and

$$W_g = \{v \in V \,|\, b(v,q) = \langle G, q\rangle \text{ for all } q \in Q\} = g + W_0$$

with $W_0 = W = \operatorname{Ker} B$, which is equivalent to the variational problem:

Find $u \in W_g$, such that

$$a(u,v) = \langle F, v\rangle \quad \text{for all } v \in W_0.$$

2.3.1 Incompressible and Almost Incompressible Materials

First we consider the variational problem for incompressible materials. For simplicity only the homogenous case $u_D = 0$ is discussed:

Find $u \in V$ and $p \in L^2(\Omega)$, such that

$$\begin{array}{lcll} a(u,v) + b(v,p) & = & \langle F, v\rangle & \text{for all } v \in V, \\ b(u,q) & = & 0 & \text{for all } q \in L^2(\Omega), \end{array}$$

with

$$a(u,v) = 2\mu \int_\Omega \varepsilon(u) : \varepsilon(v)\ dx, \quad b(v,q) = \int_\Omega q\ \mathrm{div}\, v\ dx$$

and

$$\langle F, v\rangle = \int_\Omega f \cdot v\ dx + \int_{\Gamma_N} t_N \cdot v\ ds.$$

and

$$V = \begin{cases} H^1_{0,D}(\Omega, \mathbb{R}^3) & \text{for non-trivial } \Gamma_D \\ \hat{H}(\Omega) & \text{for } \Gamma_N = \Gamma \end{cases}$$

First we consider only the case of pure Dirichlet boundary conditions ($\Gamma_D = \Gamma$):
It is obvious that p is not uniquely determined: Since

$$b(v,1) = \int_\Omega \mathrm{div}\, v\ dx = \int_\Gamma v \cdot n\ ds = 0$$

$(u, p + c)$ is also a solution for each constant $c \in \mathbb{R}$, if (u,p) is a solution. In order to guarantee uniqueness, an additional scaling condition is introduced:

$$\int_\Omega p\ dx = 0.$$

This leads to the variational problem:

Find $u \in V = H^1_0(\Omega, \mathbb{R}^3)$ and $p \in Q = L^2_0(\Omega)$, such that

$$\begin{array}{lcll} a(u,v) + b(v,p) & = & \langle F, v\rangle & \text{for all } v \in V, \\ b(u,q) & = & 0 & \text{for all } q \in Q \end{array}$$

with

$$L^2_0(\Omega) = \{q \in L^2(\Omega) : \int_\Omega q\ dx = 0\}.$$

The two variational problems are equivalent in the following sense: Each solution of this variational problem is also a solution of the original variational problem, and each solution (u,p) of the original variational problem induces a solution (u,p') with

$$p'(x) = p(x) - \frac{1}{|\Omega|} \int_\Omega p(y)\ dy.$$

From the first Korn inequality it follows that a is coercive on V and, therefore, also coercive on $\operatorname{Ker} B \subset V$.

It remains to prove the inf-sup condition.

Let $p \in L^2(\Omega)$. The gradient of p can be introduced as the linear functional $\operatorname{grad} p\colon H_0^1(\Omega, \mathbb{R}^3) \longrightarrow \mathbb{R}$, given by

$$\langle \operatorname{grad} p, v\rangle = -(p, \operatorname{div} v)_0 = -\int_\Omega p \operatorname{div} v \; dx.$$

It is easy to see that $\operatorname{grad} p \in [H_0^1(\Omega, \mathbb{R}^d)]^* = H^{-1}(\Omega, \mathbb{R}^d)$. If p is interpreted as linear functional $(p, .)_0$, then $p \in [H_0^1(\Omega)]^* = H^{-1}(\Omega)$. The norms of p in $H^{-1}(\Omega)$ and $\operatorname{grad} p$ in $H^{-1}(\Omega)$ and $H^{-1}(\Omega, \mathbb{R}^d)$ are given by

$$\|p\|_{-1} = \sup_{0\neq q\in H_0^1(\Omega)} \frac{(p,q)_0}{\|q\|_1}, \quad \|\operatorname{grad} p\|_{-1} = \sup_{0\neq v\in H_0^1(\Omega,\mathbb{R}^3)} \frac{\langle \operatorname{grad} p, v\rangle}{\|v\|_1} = \sup_{v\in H_0^1(\Omega,\mathbb{R}^3)} \frac{-(p, \operatorname{div} v)_0}{\|v\|_1}.$$

Then we have the following important inequality:

Lemma 2.4 (Nečas)**.** *Let $\Omega \subset \mathbb{R}^d$ be a bounded and open set with Lipschitz-continuous boundary. Then there exists a constant $c_N > 0$, such that*

$$\|p\|_0 \leq c_N \left(\|p\|_{-1} + \|\operatorname{grad} p\|_{-1}\right) \quad \textit{for all } p \in L^2(\Omega). \tag{2.5}$$

Proof. See [9], under stronger assumptions also [4]. □

Then it follows:

Theorem 2.6. *Let $\Omega \subset \mathbb{R}^d$ be a bounded, connected and open subset with Lipschitz-continuous boundary. Then there exists a constant $c > 0$, such that*

$$\|p\|_0 \leq c \|\operatorname{grad} p\|_{-1} \quad \textit{for all } p \in L_0^2(\Omega). \tag{2.6}$$

Proof. The embedding $i\colon H_0^1(\Omega) \longrightarrow L^2(\Omega)$ is compact. Therefore, the (adjoint) embedding $i^*\colon L^2(\Omega) \longrightarrow H^{-1}(\Omega)$ is also compact.

Assume the inequality (2.6) is not valid. There there exists a sequence (p_k) in $L_0^2(\Omega)$ with $\|p_k\|_0 = 1$ and $\|\operatorname{grad} p_k\|_{-1} \to 0$. Because of the compact embedding of $L^2(\Omega)$ in $H^{-1}(\Omega)$ there exists a convergent sub-sequence (p'_k) in $H^{-1}(\Omega)$. From (2.5) it follows that (p'_k) is a Cauchy-sequence in $L^2(\Omega)$ and, therefore, $p'_k \to p$ in $L^2(\Omega)$ with $p \in L_0^2(\Omega)$.

We have: $\operatorname{grad} p = \lim_{k\to\infty} \operatorname{grad} p'_k = 0$. Hence, p is constant, since $p \in L_0^2(\Omega)$ it follows $p = 0$, in contradiction to $\|p\|_0 = \|p_k\|_0 = 1$. □

Therefore, the inf-sup condition holds:

$$\sup_{v\in H_0^1(\Omega,\mathbb{R}^3)} \frac{(p, \operatorname{div} v)_0}{\|v\|_1} = \sup_{v\in H_0^1(\Omega,\mathbb{R}^3)} \frac{-(p, \operatorname{div} v)_0}{\|v\|_1} \geq \frac{1}{c} \|p\|_0 \quad \text{for all } p \in L_0^2(\Omega).$$

Remark: If Corollary 2.3 is applied to the case

$$X = Q = L^2_0(\Omega), \quad Y = V = H^1_0(\Omega, \mathbb{R}^3), \quad \mathcal{A} = B^* = -\operatorname{grad}\colon L^2_0(\Omega) \longrightarrow H^{-1}(\Omega, \mathbb{R}^3)$$

then the adjoint operator is the divergence:

$$\mathcal{A}^* = B = \operatorname{div}\colon H^1_0(\Omega, \mathbb{R}^3) \longrightarrow L^2_0(\Omega).$$

Condition (3"), which is equivalent to the inf-sup condition, then reads: $\operatorname{div}\colon H^1_0(\Omega, \mathbb{R}^3) \longrightarrow L^2_0(\Omega)$ is surjective, i.e.: for each $q \in L^2_0(\Omega)$ there is a $v \in H^1_0(\Omega, \mathbb{R}^3)$ with $\operatorname{div} v = q$.

The inf-sup condition can be shown in a similar way

1. for the spaces $V = H^1_{0,D}(\Omega, \mathbb{R}^3)$ and $Q = L^2(\Omega)$ in the case of non-trivial sets Γ_D and Γ_N and
2. for the spaces $V = \hat{H}(\Omega)$ and $Q = L^2(\Omega)$ in the case of pure Neumann boundary conditions ($\Gamma_N = \Gamma$).

Summary:

The inf-sup condition is equivalent to the surjectivity of the operator $B = \operatorname{div}$. Surjectivity of div is guaranteed for the following settings:

1. $\operatorname{div} : H^1_0(\Omega, \mathbb{R}^3) \longrightarrow L^2_0(\Omega)$. This covers the case $\Gamma_D = \Gamma$ (pure Dirichlet boundary conditions);
2. $\operatorname{div} : H^1_{0,D}(\Omega, \mathbb{R}^3) \longrightarrow L^2(\Omega)$ if both Γ_D and Γ_N are non-trivial (mixed boundary conditions);
3. $\operatorname{div} : \hat{H}(\Omega) \longrightarrow L^2(\Omega)$. This covers the case $\Gamma_N = \Gamma$ (pure Neumann boundary conditions).

The mixed variational problem for almost incompressible materials is equivalent to the pure displacement problem. Therefore, existence and uniqueness follow from the theorem of Lax-Milgram. However, for the condition number one obtains

$$\frac{\mu_2}{\mu_1} = \kappa(C)\frac{1}{c_K^2},$$

which approaches to infinity for $\nu \to 1/2$. The question now is whether ν-independent estimates are possible.

By the theorem of Brezzi the operator $\mathcal{K}_t : V \times Q \longrightarrow (V \times Q)^*$, given by

$$\langle \mathcal{K}_t(u,p), (v,q) \rangle = \mathcal{B}_t((u,p),(v,q)) = a(u,v) + b(v,p) + b(u,q) - t^2 c(p,q),$$

is an isomorphism for $t = 0$. Under the assumption that $c : Q \times Q \longrightarrow \mathbb{R}$ is bounded, i.e.: there is a constant γ_2 such that

$$|c(p,q)| \leq \gamma_2 \, \|p\|_Q \|q\|_Q \quad \text{for all } \lambda, \mu \in Q,$$

it immediately follows that $\mathcal{K}_t$ as a small perturbation of the isomorphism $\mathcal{K}_o$ remains an isomorphism for sufficiently small parameters t.

Under slightly stronger conditions one can show the following slightly stronger result as an extension of the theorem of Brezzi:

Theorem 2.7. *Let V, Q be real Hilbert spaces, $F \in V^*$, $G \in Q^*$, $a\colon V \times V \longrightarrow \mathbb{R}$, $b\colon V \times Q \longrightarrow \mathbb{R}$ and $c\colon Q \times Q \longrightarrow \mathbb{R}$ be bilinear forms. Assume that there exist constants α_1, α_2, β_1, β_2, $\gamma_2 > 0$ with*

1. *$|a(u,v)| \leq \alpha_2 \|u\|_V \|v\|_V$ for all $u, v \in V$,*
2. *$|b(v,q)| \leq \beta_2 \|v\|_V \|q\|_Q$ for all $v \in V$, $q \in Q$,*
3. *(a) $a(v,v) \geq 0$ for all $v \in V$,*
   (b) $a(v,v) \geq \alpha_1 \|v\|_V^2$ for all $v \in W = \operatorname{Ker} B$,
4. $\displaystyle\inf_{0 \neq q \in Q} \sup_{0 \neq v \in V} \frac{b(v,q)}{\|v\|_V \ \|q\|_Q} \geq \beta_1 > 0.$
5. *(a) $c(q,q) \geq 0$ for all $q \in Q$.*
   (b) $c(p,q) = c(q,p)$ for all $p, q \in Q$.
   (c) $|c(p,q)| \leq \gamma_2 \, \|p\|_Q \|q\|_Q$ for all $p, q \in Q$.

Then the linear operator $\mathcal{K}_t\colon V \times Q \longrightarrow (V \times Q)^*$, given by

$$\langle \mathcal{K}_t(u,p), (v,q) \rangle = \mathcal{B}_t((u,p),(v,q)),$$

with the bilinear form

$$\mathcal{B}_t((u,p),(v,q)) = a(u,v) + b(v,p) + b(u,q) - t^2 c(p,q)$$

is an isomorphism and we have

$$\|\mathcal{K}_t\| \leq \nu_2, \quad \|\mathcal{K}_t^{-1}\| \leq \frac{1}{\nu_1}$$

uniformly for all $t \in [0,1]$.

Proof. The upper bound easily follows from the boundedness of the bilinear forms:

$$\sup_{(v,q)\in V\times Q} \frac{a(u,v)+b(v,p)+b(u,q)-t^2\,c(p,q)}{\|(v,q)\|_{V\times Q}} \le \nu_2\,\|(u,p)\|_{X\times M}$$

with

$$\nu_2 = (\alpha_2^2 + 2\,\beta_2^2 + \gamma_2^2)^{1/2}.$$

For $t = 0$ it follows from the theorem of Brezzi, that there exists a constant $\mu_1 > 0$ such that

$$\sup_{(v,q)\in V\times Q} \frac{a(u,v)+b(v,p)+b(u,q)}{\|(v,q)\|_{V\times Q}} \ge \mu_1\,\|(u,p)\|_{V\times Q}$$

with

$$\|(v,q)\|_{V\times Q}^2 = \|v\|_V^2 + \|q\|_Q^2.$$

Now we have:

$$\begin{aligned} t^2\,|c(p,q)| &\le t\,|c(p,q)| \le \left[t^2\,c(p,p)\,c(q,q)\right]^{1/2} \le \gamma_2\,\left[t^2\,c(p,p)\right]^{1/2}\,\|q\|_Q \\ &\le \gamma_2\,\left[t^2\,c(p,p)\right]^{1/2}\,\|(v,q)\|_{V\times Q}. \end{aligned}$$

Hence one obtains:

$$\sup_{(v,q)\in V\times Q} \frac{a(u,v)+b(v,p)+b(u,q)-t^2c(p,q)}{\|(v,q)\|_{V\times Q}} \ge \mu_1\,\|(u,p)\|_{V\times Q} - \gamma_2\,\left[t^2\,c(p,p)\right]^{1/2}.$$

On the other side it follows (set $(v,q) = (u,-p)$):

$$\begin{aligned} &\sup_{(v,q)\in V\times Q} \frac{a(u,v)+b(v,p)+b(u,q)-t^2c(p,q)}{\|(v,q)\|_{V\times Q}} \\ &\quad\ge \frac{a(u,u)+b(u,p)+b(u,-p)-t^2c(p,-p)}{\|(u,-p)\|_{V\times Q}} = \frac{a(u,u)+t^2c(p,p)}{\|(u,p)\|_{V\times Q}} \ge \frac{t^2c(p,p)}{\|(u,p)\|_{V\times Q}} \end{aligned}$$

So, in summary, we have:

$$\begin{aligned} &\sup_{(v,q)\in V\times Q} \frac{a(u,v)+b(v,p)+b(u,q)-t^2c(p,p)}{\|(v,q)\|_{V\times Q}} \\ &\quad\ge \max\left(\mu_1\,\|(u,p)\|_{X\times_Q} - \gamma_2\,\left[t^2\,c(p,p)\right]^{1/2}, \frac{t^2c(p,p)}{\|(u,p)\|_{V\times Q}}\right). \end{aligned}$$

Since

$$\min_{y\ge 0} \max\left(\mu_1\,x - \gamma_2\,y, \frac{y^2}{x}\right) = \frac{\bar{y}^2}{x}$$

with

$$\mu_1\,x - \gamma_2\,\bar{y} = \frac{\bar{y}^2}{x},$$

i.e.:

$$\bar{y} = \left(-\frac{\gamma_2}{2} + \sqrt{\frac{\gamma_2^2}{4} + \mu_1} \right) x,$$

it follows that

$$\max \left(\mu_1 \, x - \gamma_2 \, y, \frac{y^2}{x} \right) \geq \left(-\frac{\gamma_2}{2} + \sqrt{\frac{\gamma_2^2}{4} + \mu_1} \right)^2 x$$

This implies:

$$\sup_{(v,\mu) \in V \times Q} \frac{a(u,v) + b(v,p) + b(u,\mu) - t^2 c(p,\mu)}{\|(v,\mu)\|_{V \times Q}} \geq \nu_1 \, \|(u,p)\|_{V \times Q}.$$

with

$$\nu_1 = \left(-\frac{\gamma_2}{2} + \sqrt{\frac{\gamma_2^2}{4} + \mu_1} \right)^2 = \left(\frac{2\,\mu_1}{\gamma_2 + \sqrt{\gamma_2^2 + 4\,\mu_1}} \right)^2.$$

From the symmetry of $\mathcal{B}_t((u,p),(v,\mu))$ the third condition of the theorem of Babuška-Aziz is satisfied. □

2.3.2 The Stokes Problem in Fluid Mechanics

The analysis is completely analogous to the case of incompressible materials.

2.3.3 The Hellinger-Reissner Formulation

The first variational formulation for the case of a non-trivial boundary part Γ_D with $u_D = 0$ has the following form:

Find $\sigma \in V = L^2(\Omega, \mathbb{S})$ and $u \in Q = H^1_{0,D}(\Omega, \mathbb{R}^3)$, such that

$$\begin{array}{lcll} a(\sigma,\tau) + b(\tau,u) & = & 0 & \text{for all } \tau \in V, \\ b(\sigma,v) & = & \langle G, v \rangle & \text{for all } v \in Q \end{array}$$

with

$$a(\sigma,\tau) = \int_\Omega C^{-1}\sigma : \tau \; dx, \quad b(\tau,u) = -\int_\Omega \tau : \varepsilon(u) \; dx$$

and

$$\langle G, v \rangle = -\int_\Omega f \cdot v \; dx - \int_{\Gamma_N} t_N \cdot v \; ds.$$

Obviously G is linear and bounded.
a and b are bilinear and we have:

1. a is bounded:

$$|a(\sigma,\tau)| = |(C^{-1}\sigma,\tau)_0| \leq \lambda_{\max}(C^{-1})\|\sigma\|_0\,\|\tau\|_0 = \frac{1}{\lambda_{\min}(C)}\|\sigma\|_0\,\|\tau\|_0.$$

2. b is bounded:

$$|b(\sigma,v)| = |(\sigma,\varepsilon(v))_0| \leq \|\sigma\|_0\,\|\varepsilon(v)\|_0 \leq \|\sigma\|_0\,|v|_1.$$

3. a is coercive on $\operatorname{Ker} B$, since a is coercive even on V:

$$a(\tau,\tau) = (C^{-1}\tau,\tau)_0 \geq \lambda_{\min}(C^{-1})(\tau,\tau)_0 = \frac{1}{\lambda_{\max}(C)}\|\tau\|_0^2.$$

4. b satisfies the inf-sup condition: Under the assumptions of Corollary 2.1 it follows:

$$\sup_{\tau\in L^2(\Omega,\mathbb{S})}\frac{b(\tau,v)}{\|\tau\|_0} = \sup_{\tau\in L^2(\Omega,\mathbb{S})}\frac{(\tau,\varepsilon(v))_0}{\|\tau\|_0} \geq \frac{(\varepsilon(v),\varepsilon(v))_0}{\|\varepsilon(v)\|_0} = \|\varepsilon(v)\|_0 \geq c_K\,|v|_1.$$

The second variational formulation for the case of a non-trivial boundary part Γ_N with $t_N = 0$ has the following form:

Find $\sigma \in V = H_{0,N}(\operatorname{div},\Omega,\mathbb{S})$ and $u \in Q = L^2(\Omega,\mathbb{R}^3)$, such that

$$\begin{aligned} a(\sigma,\tau) + b(\tau,u) &= \langle F,\tau\rangle \quad \text{for all } \tau \in V,\\ b(\sigma,v) &= \langle G,v\rangle \quad \text{for all } v \in Q \end{aligned}$$

with

$$a(\sigma,\tau) = \int_\Omega C^{-1}\sigma : \tau\ dx, \quad b(\tau,u) = \int_\Omega \operatorname{div}\tau\cdot u\ dx$$

and

$$\langle F,\tau\rangle = \int_{\Gamma_D}\tau\, n\cdot u_D\ ds, \quad \langle G,v\rangle = -\int_\Omega f\cdot v\ dx.$$

Obviously the functionals F and G are linear and bounded.
a and b are bilinear and we have:

1. a is bounded:

$$|a(\sigma,\tau)| \leq \frac{1}{\lambda_{\min}(C)}\|\sigma\|_0\,\|\tau\|_0 \leq \frac{1}{\lambda_{\min}(C)}\|\sigma\|_{H(\operatorname{div},\Omega,\mathbb{S})}\,\|\tau\|_{H(\operatorname{div},\Omega,\mathbb{S})}.$$

2. b is bounded:

$$|b(\sigma,v)| = |(\operatorname{div}\sigma,v)_0\| \leq \|\operatorname{div}\sigma\|_0\,\|v\|_0 \leq \|\sigma\|_{H(\operatorname{div},\Omega,\mathbb{S})}\,\|v\|_0.$$

3. a is coercive on $\operatorname{Ker} B$:

$$\begin{aligned}\operatorname{Ker} B &= \{\tau \in H_{0,N}(\operatorname{div}, \Omega, \mathbb{S}) \,|\, (\operatorname{div}\tau, v)_0 = 0 \text{ for all } v \in L^2(\Omega, \mathbb{R}^3)\} \\ &= \{\tau \in H_{0,N}(\operatorname{div}, \Omega, \mathbb{S}) \,|\, \operatorname{div}\tau = 0\}.\end{aligned}$$

Hence

$$a(\tau,\tau) \geq \frac{1}{\lambda_{\max}(C)}\|\tau\|_0^2 = \frac{1}{\lambda_{\max}(C)}\|\tau\|_{H(\operatorname{div},\Omega,\mathbb{S})}^2 \quad \text{for all } \tau \in \operatorname{Ker} B.$$

4. b satisfies the inf-sup condition: One has to show that $\operatorname{div}\colon H_{0,N}(\operatorname{div}, \Omega, \mathbb{S}) \longrightarrow L^2(\Omega, \mathbb{R}^3)$ is surjective: Let $v \in L^2(\Omega, \mathbb{R}^3)$ be given. There exists a $\tau \in H_{0,N}(\operatorname{div}, \Omega, \mathbb{S})$ with $\operatorname{div}\tau = v$. For this we choose the ansatz $\tau = \varepsilon(u)$ with

$$(\varepsilon(u), \varepsilon(w))_0 = -(v, w)_0 \quad \text{for all } w \in H^1_{0,D}(\Omega, \mathbb{R}^3).$$

From the discussion of the primal variational formulation the existence of such a u and, consequently, the existence of τ is guaranteed and the following estimates hold:

$$\|\varepsilon(u)\|_0^2 \leq \|v\|_0 \, \|u\|_0 \leq c_F \, \|v\|_0 \, |u|_1 \leq \frac{c_F}{c_K} \, \|v\|_0 \, \|\varepsilon(u)\|_0,$$

hence

$$\|\tau\|_0 \leq \frac{c_F}{c_K} \, \|v\|_0$$

and, therefore,

$$\|\tau\|_{H(\operatorname{div},\Omega,\mathbb{S})}^2 \leq \left(1 + \frac{c_F^2}{c_K^2}\right) \, \|v\|_0^2$$

This finally implies

$$\sup_{\tau \in V_0} \frac{(\operatorname{div}\tau, v)_0}{\|\tau\|_{H(\operatorname{div},\Omega,\mathbb{S})}} \geq \frac{c_K}{\sqrt{c_F^2 + c_K^2}} \, \|v\|_0.$$

So far the analysis was based on the coercivity in $L^2(\Omega, \mathbb{S})$

$$a(\tau,\tau) \geq \frac{1}{\lambda_{\max}(C)} \, \|\tau\|_0^2 \quad \text{for all } \tau \in L^2(\Omega, \mathbb{S})$$

for the bilinear form

$$a(\sigma,\tau) = \int_\Omega C^{-1}\sigma : \tau \, dx = (C^{-1}\sigma, \tau)_0.$$

For St.Venant-Kirchhoff materials in the almost incompressible case it follows that $\nu \longrightarrow 1/2$ and $1/\lambda_{\max}(C) \longrightarrow 0$, while the norm of a is bounded uniformly in ν. This implies that the estimate of the condition number of the problem approaches infinity in this case.

By a refined analysis of the Hellinger-Reissner formulation one can actually show ν-independent estimates.

Actually, by the theorem of Brezzi the $L^2(\Omega,\mathbb{S})$-coercivity is needed only on the subspace

$$\begin{aligned} Z &= \left\{\tau \in L^2(\Omega,\mathbb{S}) \,\middle|\, (\tau,\varepsilon(v))_0 = 0 \text{ for all } v \in H^1_{0,D}(\Omega,\mathbb{R}^3)\right\} \\ &= \{\tau \in H_{0,N}(\mathrm{div},\Omega,\mathbb{S}) \mid \mathrm{div}\,\tau = 0\} \end{aligned}$$

We have

Lemma 2.5. *There exists a constant $c > 0$ independent of ν with*

$$\int_\Omega C^{-1}\tau : \tau \; dx \geq c \; \|\tau\|_0^2 \quad \textit{for all } \tau \in Z.$$

Proof. We have:

$$\sigma(C^{-1}) = \left\{\frac{1-2\nu}{E}, \frac{1+\nu}{E}\right\}$$

$\lambda_1 = (1-2\nu)/E$ is a simple eigenvalue of C^{-1} with eigenvector $\sigma_1 = I$ and $\lambda_2 = (1+\nu)/E$ is an eigenvalue of C^{-1} with multiplicity 8.

An arbitrary element $\tau \in Z$ can be written in the following way:

$$\tau = \tau_1 + \tau_D \quad \text{with } \tau_1 = \frac{1}{3}\,\mathrm{trace}(\tau)\,I \text{ and } \tau_D = \tau - \frac{1}{3}\,\mathrm{trace}(\tau)\,I.$$

Since $(\tau_D, I)_0 = 0$, it follows that:

$$(C^{-1}\tau,\tau)_0 = \lambda_1(\tau_1,\tau_1)_0 + \lambda_2(\tau_D,\tau_D)_0.$$

From the inf-sup condition for div it follows that there exists an element $v \in H^1_{0,D}(\Omega,\mathbb{R}^3)$ such that

$$\mathrm{div}\, v = \mathrm{trace}(\tau) \quad \text{with} \quad \|v\|_1 \leq c_1 \,\|\,\mathrm{trace}(\tau)\|_0.$$

Then

$$\begin{aligned} (\tau_1,\tau_1)_0 &= \frac{1}{3}(\mathrm{trace}(\tau),\mathrm{trace}(\tau))_0 = \frac{1}{3}(\mathrm{div}\, v,\mathrm{trace}(\tau))_0 = \frac{1}{3}(\varepsilon(v),\mathrm{trace}(\tau)\,I)_0 \\ &= (\varepsilon(v),\tau_1)_0 = (\varepsilon(v),\tau-\tau_D)_0 = (v,\mathrm{div}\,\tau)_0 - (\varepsilon(v),\tau_D)_0 = -(\varepsilon(v),\tau_D)_0, \end{aligned}$$

This implies the following estimates:

$$\|\tau_1\|_0^2 \leq \|\varepsilon(v)\|_0\,\|\tau_D\|_0 \leq \|v\|_1\,\|\tau_D\|_0 \leq c_1\,\|\,\mathrm{trace}(\tau)\|_0\,\|\tau_D\|_0 = \sqrt{3}\,c_1\,\|\tau_1\|_0\,\|\tau_D\|_0,$$

hence

$$\|\tau_1\|_0 \leq \sqrt{3}\,c_1\,\|\tau_D\|_0$$

and, therefore,

$$\|\tau\|_0^2 = \|\tau_1\|_0^2 + \|\tau_D\|_0^2 \leq (1+3\,c_1^2)\|\tau_D\|_0^2.$$

This implies:

$$(C^{-1}\tau,\tau)_0 \geq \lambda_2(\tau_D,\tau_D)_0 \geq \frac{\lambda_2}{1+3\,c_1^2}\,\|\tau_D\|_0^2 \geq \frac{1}{E(1+3\,c_1^2)}\,\|\tau\|_0^2.$$

□

Chapter 3

Finite Element Methods

3.1 FEM for the Primal Variational Problem

The pure displacement problem in elastostatics leads (after homogenization) to a (primal) variational problem of the following form:

Find $u \in V$, such that

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

with $V \subset H^1(\Omega, \mathbb{R}^3)$.

We use Galerkin's principle for discretization: An appropriate finite-dimensional subspace $V_h \subset V$ is chosen and an approximate solution $u_h \in V_h$ is computed as the solution of the (finite-dimensional) variational problem:

$$a(u_h, v_h) = \langle F, v_h \rangle \quad \text{for all } v_h \in V_h.$$

The standard assumptions of the theorem of Lax-Milgram for the continuous problem have been shown, therefore, the standard assumptions of the theorem of Lax-Milgram are also satisfied for the discrete problem. Hence there exists a unique solution of the discrete problem and the solution depends continuously on the data.

Under the standard assumptions of the theorem of Lax-Milgram Cea's lemma gives the following estimate for the discretization error:

$$\|u - u_h\|_V \leq \frac{\mu_2}{\mu_1} \inf_{v_h \in V_h} \|u - v_h\|_V.$$

So the discretization error can be estimates by the approximation error. The spaces V_h are to be chosen such that the functions in V can be accurately approximated by functions in V_h. The finite element method is based on a subdivision of the domain $\Omega \subset \mathbb{R}^3$ in polyhedra (e.g.: tetrahedra, hexahedra, ...). The functions in V_h are typically piecewise polynomial functions with respect to this subdivision. In order to obtain conforming function spaces, i.e. $V_h \subset V \subset H^1(\Omega, \mathbb{R}^3)$, the functions have to be continuous.

A few examples of C^0-elements (continuous elements):

1. The P_1-element on a tetrahedral subdivision: For each component of the displacement continuous and piecewise linear elements are used.

2. The Q_1-element on a hexahedral subdivision: For each component of the displacement on the unit cube (the reference element) trilinear elements (i.e. linear with respect to each coordinate) are used. By a trilinear transformation from the unit cube to an arbitrary hexahedron the so-called isoparametric trilinear element on hexahedral subdivisions results.

3. Higher order elements on tetrahedral subdivisions (P_k-elements, continuous and piecewise polynomial of degree $\leq k$) or on hexahedral subdivisions (Q_k-elements, piecewise polynomial of degree $\leq k$ in each coordinate of the unit cube, transformation to arbitrary hexahedra).

Under appropriate assumptions the approximation error of P_k- and Q_k-elements can be estimated by:

$$\inf_{v_h \in V_h} \|u - v_h\|_1 \leq c\, h^k \, \|u\|_{k+1}.$$

For the actual computation of the approximate solution $u_h \in V_h$ we need a basis $\{\varphi_i \in V_h : i = 1, \ldots, N\}$ of V_h. Then each function $u_h \in V_h$ can be represented in the form

$$u_h(x) = \sum_{j=1}^{N} u_j \, \varphi_j(x).$$

For the vector $\underline{u}_h$ of the coefficients a linear system of equations results from the discrete variational problem:

$$K_h \underline{u}_h = \underline{f}_h.$$

The so-called stiffness matrix K_h is symmetric and positive definite as a consequence of the properties of the bilinear form a.

The condition number of the stiffness matrix K_h is a measure of the degree of difficulty for solving the linear system. Typically we have

$$\kappa(K_h) = \frac{\mu_2}{\mu_1} \, O(h^{-2}),$$

where h denotes the mesh size of the subdivision (e.g.: the length of the longest edge of a tetrahedral or hexahedral subdivision).

Efficient methods for solving the linear systems are multilevel or multigrid methods. These methods can be accelerated by Krylov subspace methods (e.g. the CG method).

This short review shows the importance of the condition number μ_2/μ_1 of the variational problem the discretization error as well as for the solution methods of the linear system.

In the case of almost incompressible materials the condition number μ_2/μ_1 diverges to ∞. This leads to a large discretization error and to growing difficulties for solving the linear systems. The actually computed displacements u are too small, in general (locking). A remedy of this problems is provided by FE methods which are based on the mixed variational formulation.

3.2 Mixed Finite Element Methods

An approximate solution of the mixed variational problem

$$\begin{array}{rcll} a(u,v)+b(v,p) & = & \langle F,v\rangle & \text{for all } v\in V \\ b(u,q) & = & \langle G,q\rangle & \text{for all } q\in Q \end{array} \qquad (P)$$

is obtained by chosen appropriate finite-dimensional subspaces

$$V_h\subset V,\quad Q_h\subset Q.$$

By Galerkin's principle the approximate solutions $u_h\in V_h$ and $p_h\in Q_h$, solve the discrete variational problem

$$\begin{array}{rcll} a(u_h,v_h)+b(v_h,p_h) & = & \langle F,v_h\rangle & \text{for all } v_h\in V_h \\ b(u_h,q_h) & = & \langle G,q_h\rangle & \text{for all } q_h\in Q_h. \end{array} \qquad (P_h)$$

The analysis of the discrete problem (P_h) is done analogously to the problem (P). We have the following generalization of Cea's lemma:

Theorem 3.1. *Assume the notations and assumptions of the theorem of Brezzi (2.5). Let $V_h\subset V$, $Q_h\subset Q$ be finite-dimensional subspaces. Assume that there exist constants $\widetilde{\alpha}_1$, $\widetilde{\beta}_1$ with*

3' $a(v_h,v_h)\geq\widetilde{\alpha}_1\|v_h\|_V^2$ for all $v_h\in W_h=\operatorname{Ker}B_h=\{v_h\in V_h\colon b(v_h,q_h)=0$ for all $q_h\in Q_h\}$,

4' $\inf\limits_{0\neq q_h\in Q_h}\sup\limits_{0\neq v_h\in V_h}\dfrac{b(v_h,q_h)}{\|v_h\|_V\,\|q_h\|_Q}\geq\widetilde{\beta}_1>0.$

Then the problem (P_h) has a unique solution $(u_h,p_h)\in V_h\times Q_h$ and:

$$\begin{aligned} \|u-u_h\|_V &\leq \left(1+\frac{\alpha_2}{\widetilde{\alpha}_1}\right)\left(1+\frac{\beta_2}{\widetilde{\beta}_1}\right)\inf_{v_h\in V_h}\|u-v_h\|_V+\frac{\beta_2}{\widetilde{\alpha}_1}\inf_{q_h\in Q_h}\|p-q_h\|_Q, \\ \|p-p_h\|_Q &\leq \left(1+\frac{\alpha_2}{\widetilde{\alpha}_1}\right)\left(1+\frac{\beta_2}{\widetilde{\beta}_1}\right)\frac{\alpha_2}{\widetilde{\beta}_1}\inf_{v_h\in V_h}\|u-v_h\|_V \\ &\quad+\left[1+\frac{\beta_2}{\widetilde{\beta}_1}\left(1+\frac{\alpha_2}{\widetilde{\alpha}_1}\right)\right]\inf_{q_h\in Q_h}\|p-q_h\|_Q. \end{aligned}$$

Proof. The existence and uniqueness of $u_h\in V_h$ and $p_h\in Q_h$ follows from the theorem of Brezzi. We have

$$\begin{array}{rcll} a(u,w)+b(w,p) & = & \langle F,w\rangle & \text{for all } w\in V, \\ b(u,r) & = & \langle G,r\rangle & \text{for all } r\in Q. \end{array}$$

and

$$\begin{array}{rcll} a(u_h,w_h)+b(w_h,p_h) & = & \langle F,w_h\rangle & \text{for all } w_h\in V_h, \\ b(u_h,r_h) & = & \langle G,r_h\rangle & \text{for all } r_h\in Q_h. \end{array}$$

By subtracting one obtains

$$\begin{array}{lcll} a(u_h - u, w_h) + b(w_h, p_h - p) & = & 0 & \text{for all } w_h \in V_h, \\ b(u_h - u, r_h) & = & 0 & \text{for all } r_h \in Q_h. \end{array}$$

Hence, we have for arbitrary $v_h \in V_h$ and $q_h \in Q_h$

$$\begin{array}{lcll} a(u_h - v_h, w_h) + b(w_h, p_h - q_h) & = & a(u - v_h, w_h) + b(w_h, p - q_h) & \text{for all } w_h \in V_h, \\ b(u_h - v_h, r_h) & = & b(u - v_h, r_h) & \text{for all } r_h \in Q_h. \end{array}$$

From the theorem of Brezzi it follows that

$$\begin{aligned} \|u_h - v_h\|_V &\leq \frac{1}{\widetilde{\alpha}_1}\|\widetilde{F}\|_{V_h^*} + \frac{1}{\widetilde{\beta}_1}\left(1 + \frac{\alpha_2}{\widetilde{\alpha}_1}\right)\|\widetilde{G}\|_{Q_h^*} \\ \|p_h - q_h\|_Q &\leq \frac{1}{\widetilde{\beta}_1}\left(1 + \frac{\alpha_2}{\widetilde{\alpha}_1}\right)\|\widetilde{F}\|_{V_h^*} + \frac{\alpha_2}{\widetilde{\beta}_1^2}\left(1 + \frac{\alpha_2}{\widetilde{\alpha}_1}\right)\|\widetilde{G}\|_{Q_h^*} \end{aligned}$$

with

$$\langle \widetilde{F}, w_h \rangle = a(u - v_h, w_h) + b(w_h, p - q_h) \quad \text{and} \quad \langle \widetilde{G}, r_h \rangle = b(u - v_h, r_h).$$

Now we have:

$$\|\widetilde{F}\|_{V_h^*} \leq \alpha_2 \|u - v_h\|_V + \beta_2 \|p - q_h\|_Q \quad \text{and} \quad \|\widetilde{G}\|_{Q_h^*} \leq \beta_2 \|u - v_h\|_V.$$

With

$$\|u - u_h\|_V \leq \|u - v_h\|_V + \|u_h - v_h\|_V \quad \text{and} \quad \|p - p_h\|_Q \leq \|p - q_h\|_Q + \|p_h - q_h\|_Q$$

the statement easily follows. □

Observe that $\operatorname{Ker} B_h \not\subset \operatorname{Ker} B$, in general. Therefore, the coercivity of a on $\operatorname{Ker} B$ does not necessarily imply the coercivity of a on $\operatorname{Ker} B_h$.

Similarly, the continuous inf-sup condition does not necessarily imply the discrete inf-sup condition.

So the assumptions $(3')$ and $(4')$ must be explicitly verified for the chosen subspaces V_h and Q_h.

If these assumptions $(3')$ and $(4')$ hold with constants which are independent of h, then the discretization error approaches 0 for $h \to 0$ if the approximation error does so.

A very helpful tool for showing the discrete inf-sup condition

$$\inf_{0 \neq q_h \in Q_h} \sup_{0 \neq v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V \, \|q_h\|_Q} \geq \widetilde{\beta}_1 > 0, \tag{3.1}$$

is the following lemma:

Lemma 3.1. *Assume there exists a linear operator $\Pi_h \colon V \to V_h$ with*

1. $b(\Pi_h v, q_h) = b(v, q_h)$ *for all* $q_h \in Q_h$ *and all* $v \in V$ *and*

2. $\|\Pi_h v\|_V \leq c \, \|v\|_V$ *for all* $v \in V$.

Then the inf-sup condition for b *and the spaces* V *and* Q *with a constant* $\beta_1 \geq 0$ *implies the discrete inf-sup condition for* b *and the spaces* V_h *and* Q_h *with a constant* $\widetilde{\beta}_1 = \beta_1 / c$.

Proof. We have:

$$\beta_1 \, \|q_h\|_Q \leq \sup_{0 \neq v \in V} \frac{b(v, q_h)}{\|v\|_V} \leq c \sup_{0 \neq v \in V} \frac{b(\Pi_h v, q_h)}{\|\Pi_h v\|_V} \leq c \sup_{0 \neq v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_V}.$$

By dividing by c the statement follows. □

The operator Π_h is called a Fortin operator.

For the actual computations of the approximate solution $(u_h, p_h) \in V_h \times Q_h$ we need bases$\{\varphi_j\}$ for V_h and $\{\psi_k\}$ for Q_h. Then these approximate solutions can be represented in the following form:

$$u_h = \sum_j u_j \; \varphi_j, \quad p_h = \sum_k p_k \; \psi_k.$$

From the discrete variational problem a linear system of equations is obtained:

$$\begin{pmatrix} A_h & B_h^T \\ B_h & 0 \end{pmatrix} \begin{pmatrix} \underline{u}_h \\ \underline{p}_h \end{pmatrix} = \begin{pmatrix} \underline{f}_h \\ \underline{g}_h \end{pmatrix}$$

with

$$\begin{aligned} A_h &= (a(\varphi_j, \varphi_i)\,), \\ B_h &= (b(\varphi_j, \psi_k)\,), \\ \underline{u}_h &= (u_j)\,, \quad \underline{p}_h = (p_k)\,, \quad \underline{f}_h = \quad (\langle F, \varphi_i \rangle)\,, \quad \underline{g}_h = \quad (\langle G, \psi_k \rangle)\,. \end{aligned}$$

3.3 Mixed FEM for the Stokes Problem

For simplicity only the case $\Gamma_D = \Gamma$ (pure Dirichlet boundary conditions) with $u_D = 0$ is considered.

The bilinear form a is coercive on $V = H_0^1(\Omega, \mathbb{R}^d)$, therefore, a is also coercive on $\operatorname{Ker} B_h \subset V$ with the same h-independent constant $\widetilde{\alpha}_1 = \alpha_1$.

The discrete inf-sup condition for the bilinear form b with an h-independent constant $\widetilde{\beta}_1 > 0$ has to be investigated explicitly.

3.3.1 The Q_1-P_0 Element

Let $\Omega = (-1,1) \times (-1,1)$, $n \in \mathbb{N}$ and $h = 1/(2n)$. The nodes (x_i, y_j) with $x_i = i\,h$ and $y_j = j\,h$ define a subdivision

$$\mathcal{T}_h = \{T_{i,j} \Big| \; i,j = -2n, \ldots, 2n-1\}$$

of Ω with the squares $T_{i,j} = (x_i, x_{i+1}) \times (y_j, y_{j+1})$.

The following spaces are introduced:

$$C_0(\overline{\Omega}, \mathbb{R}^2) = \{v \in C(\overline{\Omega}, \mathbb{R}^2) \Big| \; v = 0 \text{ on } \Gamma\}$$

and

$$\begin{aligned} P_k &= \{w(x,y) = \sum_{0 \le i+j \le k} c_{ij}\; x^i y^j\}, \\ Q_k &= \{w(x,y) = \sum_{0 \le i,j \le k} c_{ij}\; x^i y^j\}. \end{aligned}$$

Then the spaces V_h and Q_h are defined by:

$$V_h = \{v \in C_0(\overline{\Omega}, \mathbb{R}^2) \Big| \; v|_T \in Q_1 \text{ for all } T \in \mathcal{T}_h\}$$

and

$$Q_h = \hat{Q}_h \cap L^2_0(\Omega) \quad \text{with} \quad \hat{Q}_h = \{q \in L^2(\Omega) \Big| \; q|_T \in P_0 \text{ for all } T \in \mathcal{T}_h\}.$$

Obviously we have

$$V_h \subset V = H^1_0(\Omega, \mathbb{R}^2) \quad \text{and} \quad Q_h \subset Q = L^2_0(\Omega).$$

These spaces satisfy the following approximation properties:

1. For $u \in H^1_0(\Omega, \mathbb{R}^2)$ and $p \in L^2_0(\Omega)$ we have:

$$\lim_{h \to 0} \inf_{v_h \in V_h} \|u - v_h\|_1 = 0 \quad \text{and} \quad \lim_{h \to 0} \inf_{q_h \in Q_h} \|p - q_h\|_0 = 0.$$

2. Under the stronger assumption $u \in H^1_0(\Omega, \mathbb{R}^2) \cap H^2(\Omega, \mathbb{R}^2)$ and $p \in L^2_0(\Omega) \cap H^1(\Omega)$ we have: There is a constant C with

$$\inf_{v_h \in V_h} \|u - v_h\|_1 \le C\,h\,\|u\|_2 \quad \text{and} \quad \inf_{q_h \in Q_h} \|p - q_h\|_0 \le C\,h\,\|p\|_1.$$

Discussion of the inf-sup condition:

The functions $\varphi_{i,j} \in \{v \in C_0(\overline{\Omega}) \mid v|_T \in Q_1 \text{ for all } T \in \mathcal{T}_h\}$ are given by the conditions

$$\varphi_{i,j}(x_k, y_l) = \delta_{(i,j),(k,l)}.$$

The following basis functions for V_h are used:

$$\varphi_{i,j} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \varphi_{i,j} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \quad \text{with } i,j = -2n+1, \ldots, 2n-1.$$

Then the following representation for an arbitrary function $v_h \in V_h$ follows:

$$v_h = \sum_{i,j=-2n+1}^{2n-1} \varphi_{i,j} \begin{pmatrix} u_{i,j} \\ v_{i,j} \end{pmatrix}.$$

The following basis functions for $\hat{Q}_h$ are used:

$$\psi_{i,j}(x,y) = \begin{cases} 1 \text{ for } (x,y) \in T_{i,j}, \\ 0 \text{ for } (x,y) \notin T_{i,j}. \end{cases}$$

Then the following representation for an arbitrary function $q_h \in \hat{Q}_h$ follows:

$$q_h = \sum_{i,j=-2n}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}}\, \psi_{i,j}.$$

With these representations one obtains:

$$\begin{aligned} -\int_\Omega q_h \operatorname{div} v_h \, dx &= -\sum_{i,j=-2n}^{2n-1} \int_{T_{i,j}} q_h \operatorname{div} v_h \, dx = -\sum_{i,j=-2n}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}} \int_{\partial T_{i,j}} v_h \cdot n \, ds \\ &= -\sum_{i,j=-2n}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}} \left[\frac{1}{2}(u_{i+1,j+1} + u_{i+1,j}) + \frac{1}{2}(v_{i,j+1} + v_{i+1,j+1}) \right. \\ &\qquad \left. -\frac{1}{2}(u_{i,j+1} + u_{i,j}) - \frac{1}{2}(v_{i,j} + v_{i+1,j}) \right] h. \end{aligned}$$

Since

$$\begin{aligned} \sum_{i,j=-2n}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{i,j} &= \sum_{i,j=-2n+1}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{i,j} \\ \sum_{i,j=-2n}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{i+1,j} &= \sum_{i=-2n}^{2n-2} \sum_{j=-2n+1}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{i+1,j} = \sum_{i,j=-2n+1}^{2n-1} q_{i-\frac{1}{2},j+\frac{1}{2}}\, w_{i,j} \\ \sum_{i,j=-2n}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{1,j+1} &= \sum_{i=-2n+1}^{2n-1} \sum_{j=-2n}^{2n-2} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{i,j+1} = \sum_{i,j=-2n+1}^{2n-1} q_{i+\frac{1}{2},j-\frac{1}{2}}\, w_{i,j} \\ \sum_{i,j=-2n}^{2n-1} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{i+1,j+1} &= \sum_{i,j=-2n}^{2n-2} q_{i+\frac{1}{2},j+\frac{1}{2}}\, w_{i+1,j+1} = \sum_{i,j=-2n+1}^{2n-1} q_{i-\frac{1}{2},j-\frac{1}{2}}\, w_{i,j} \end{aligned}$$

it follows that

$$-\int_\Omega q_h \operatorname{div} v_h \, dx = h^2 \sum_{i,j=-2n+1}^{2n-1} \left[u_{i,j} \left(\nabla_1 q\right)_{i,j} + v_{i,j} \left(\nabla_2 q\right)_{i,j} \right]$$

with

$$\begin{aligned} \left(\nabla_1 q\right)_{i,j} &= \frac{1}{2h} \left[q_{i+\frac{1}{2},j-\frac{1}{2}} + q_{i+\frac{1}{2},j+\frac{1}{2}} - q_{i-\frac{1}{2},j-\frac{1}{2}} - q_{i-\frac{1}{2},j+\frac{1}{2}} \right], \\ \left(\nabla_2 q\right)_{i,j} &= \frac{1}{2h} \left[q_{i-\frac{1}{2},j+\frac{1}{2}} + q_{i+\frac{1}{2},j+\frac{1}{2}} - q_{i-\frac{1}{2},j-\frac{1}{2}} - q_{i+\frac{1}{2},j-\frac{1}{2}} \right]. \end{aligned}$$

From this representation it is easy to see that the function

$$\mu_h = \sum_{i,j=-2n}^{2n-1} \mu_{i+\frac{1}{2},j+\frac{1}{2}} \, \psi_{i,j}$$

with

$$\mu_{i+\frac{1}{2},j+\frac{1}{2}} = (-1)^{i+j}$$

satisfies:

$$b(v_h, \mu_h) = -\int_\Omega \mu_h \operatorname{div} v_h \, dx = 0 \quad \text{for all } v_h \in V_h.$$

That means that $\mu_h \in \operatorname{Ker} B_h^T$. Since, additionally,

$$\int_\Omega \mu_h = 0,$$

it follows that $\mu_h \in Q_h$ and the inf-sup is not satisfied.

Remark: The function μ_h is called a "spurious pressure mode", in this particular case it is also called a "checkerboard mode" ("checkerboard instability").

A first attempt to stabilize the Q_1-P_0 element is to consider only those functions $q_h \in Q_h$ which are orthogonal to μ_h:

$$\overline{Q}_h = \{ q_h \in Q_h \Big| \int_\Omega q_h \, \mu_h \, dx = 0 \}.$$

Since the constant functions and the multiples of μ_h are the only functions in $\operatorname{Ker} B_h^T$, there exists a constant $\widetilde{\beta}_{1,h}$ with

$$\inf_{0 \neq q_h \in \overline{Q}_h} \sup_{0 \neq v_h \in V_h} \frac{b(v_h, q_h)}{\|v_h\|_1 \, \|q_h\|_0} \geq \widetilde{\beta}_{1,h} > 0.$$

However, it can be shown that

$$\widetilde{\beta}_{1,h} = O(h).$$

Hence there is no lower bound $\widetilde{\beta}_1 > 0$ which is independent of h.

In order to stabilize the Q_1-P_0 element, the space Q_h must be further reduced: By constructing one macro-element from 4 neighboring elements of the original subdivision $\mathcal{T}_h$

$$M_{i,j} = (x_{2i}, x_{2i+2}) \times (y_{2j}, y_{2j+2}), \quad i,j = -n, n-1,$$

a second subdivision is obtained:

$$\mathcal{M}_h = \{M_{i,j} \Big| \; i,j = -n, \ldots, n-1\}.$$

We introduce the following space:

$$Q_{2h} = \{q \in L_0^2(\Omega) \Big| \; q|_M \in P_0 \text{ for all } M \in \mathcal{M}_h\}.$$

Now we have

Theorem 3.2. *For the spaces V_h and Q_{2h} the discrete inf-sup condition is satisfied with a constant independent of h.*

Sketch of the proof: Let $v \in H_0^1(\Omega, \mathbb{R}^2)$. By Lemma 3.1 a $v_h \in V_h$ must be constructed such that

$$\int_\Omega q_h \; \mathrm{div}(v_h - v) \; dx = 0 \quad \text{for all } q_h \in Q_{2h}.$$

This is equivalent to

$$0 = \int_M \mathrm{div}(v_h - v) \; dx = \int_{\partial M} (v_h - v) \cdot n \; ds \quad \text{for all } M \in \mathcal{M}_h.$$

Let $M \in \mathcal{M}_h$ be an arbitrary macro-element with vertices x_1, x_2, x_3, x_4 and midpoint x_5 and the edges S_1, S_2, S_3, S_4.

From the analysis from above it suffices for v_h to satisfy the condition

$$\int_{S_i} v_h \; ds = \int_{S_i} v \; ds \quad \text{for } i = 1,2,3,4.$$

Usually a function in V_h is defined by its values at the nodes. Here we fix v_h by the following conditions:

1. For the four vertices and the midpoint of the macro we prescribe the value of v_h:

$$v_h(x_i) = \begin{cases} \dfrac{1}{|\Delta_i|} \displaystyle\int_{\Delta_i} v \; dx & \text{for } x_i \in \Omega, \\ 0 & \text{for } x_i \in \Gamma, \end{cases}$$

   where Δ_i is the union of all elements T_j of the original subdivision whose closure $\overline{T}_j$ contain x_i as a vertex.

2. Instead of prescribing the value of v_h in the midpoints of the edges, we require

$$\int_{S_i} v_h \, ds = \int_{S_i} v \, ds.$$

It is easy to check that v_h is well-defined and continuous and vanishes on the boundary $\Gamma = \partial\Omega$, therefore $v_h \in V_h$, and that the mapping $v \mapsto v_h$ is linear.

By a so-called scaling argument one shows the existence of an h-independent constant C with

$$\|v_h\|_1 \leq C\|v\|_1.$$

□

Remark:

1. Although the discrete inf-sup condition is not satisfied for the original spaces V_h and Q_h, the convergence $u_h \to u$ can be shown. However, p_h does not, in general, converge to p.

2. The results can easily be carried over to more general quadrilateral subdivisions by using the isoparametric bilinear element.

3.3.2 The P_1-P_0 Element

The corresponding element on triangular subdivisions is the P_1-P_0 element. In this case, we have, in general

$$\operatorname{Ker} B_h = \{0\},$$

since

$$\dim Q_h > \dim V_h.$$

Proof. Obviously we have:

$$\dim V_h = 2\, N_i, \quad \dim Q_h = N_e,$$

where N_i denotes the number of nodes in Ω and N_e denotes the number of triangles of the subdivision. For a general triangular subdivision we have:

$$N_e = 2\, N_i + N_r - 2 > 2\, N_i,$$

where N_r denotes the number of nodes in Γ. □

The P_1-P_0 element can be stabilized analogously to the stabilization of the Q_1-P_0 element: Let $\mathcal{T}_h$ be a triangular subdivision of Ω and let $\mathcal{T}_{h/2}$ be that refined triangular subdivision which is obtained by uniform refinement: Each triangle $T \in \mathcal{T}_h$ is subdivided into four congruent sub-triangles.

The spaces

$$V_h = \{v \in C_0(\overline{\Omega}, \mathbb{R}^2) \Big| \; v|_T \in P_1 \text{ for all } T \in \mathcal{T}_{h/2}\}$$

and

$$Q_h = \{q \in L_0^2(\Omega) \Big| \; q|_T \in P_0 \text{ for all } T \in \mathcal{T}_h\}$$

satisfy the discrete inf-sup condition with an h-independent constant. The proof is done by constructing a Fortin operator analogously to the case of the Q_1-P_0 element.

The stabilized Q_1-P_0 element and the stabilized P_1-P_0 element are suitable finite elements for the mixed variational problem for incompressible and almost incompressible materials.

In the following we discuss in more details the application of this element for almost incompressible materials:

The first equation reads

$$2\mu \int_\Omega \varepsilon(u_h) : \varepsilon(v_h) \; dx + \int_\Omega p_h \; \mathrm{div}\, v_h \; dx = \langle F, v_h \rangle.$$

Since p_h is piecewise constant, it follows

$$\int_\Omega p_h \; \mathrm{div}\, v_h \; dx = \sum_{T \in \mathcal{T}_h} \int_T p_h \; \mathrm{div}\, v_h \; dx = \sum_{T \in \mathcal{T}_h} \frac{1}{|T|} \int_T p_h \; dx \int_T \mathrm{div}\, v_h \; dx,$$

where $|T|$ denotes the area (volume) of T.

From the second equation

$$\int_\Omega q_h \; \mathrm{div}\, u_h \; dx - \frac{1}{\lambda} \int_\Omega p_h \, q_h \; dx = 0 \quad \text{for all } q_h \in \hat{Q}_h$$

we obtain:

$$\int_T p_h \; dx = \lambda \int_T \mathrm{div}\, u_h \; dx \quad \text{for all } T \in \mathcal{T}_h.$$

Hence

$$\int_\Omega p_h \; \mathrm{div}\, v_h \; dx = \lambda \sum_{T \in \mathcal{T}_h} \frac{1}{|T|} \int_T \mathrm{div}\, u_h \; dx \int_T \mathrm{div}\, v_h \; dx = \lambda \sum_{T \in \mathcal{T}_h} \overline{\mathrm{div}\, u_h}^T \, \overline{\mathrm{div}\, v_h}^T \, |T|,$$

where $\overline{\mathrm{div}\, w_h}^T$ denotes the mean value of $\mathrm{div}\, w_h$ over the element T. In summary, the following (primal) variational problem for u_h results:

$$2\mu \int_\Omega \varepsilon(u_h) : \varepsilon(v_h) \; dx + \lambda \sum_{T \in \mathcal{T}_h} \overline{\mathrm{div}\, u_h}^T \, \overline{\mathrm{div}\, v_h}^T \, |T| = \langle F, v_h \rangle \quad \text{for all } v_h \in V_h.$$

This coincides with the discretization of the pure displacement problem in V_h, except that instead of the original second term

$$\lambda \int_\Omega \mathrm{div}\, u_h \; \mathrm{div}\, v_h \; dx$$

the following approximation by a quadrature rule is used:

$$\lambda \sum_{T\in\mathcal{T}_h} \overline{\operatorname{div} u_h}^T \, \overline{\operatorname{div} v_h}^T \, |T|.$$

This technique is called selective reduced integration.

3.3.3 The MINI Element

Let $\mathcal{T}_h$ be a triangular subdivision of the domain $\Omega \subset \mathbb{R}^2$. Consider the so-called P_1-P_1 element on this subdivision, given by

$$V_h = \{v \in C_0(\overline{\Omega}, \mathbb{R}^2) : \ v|_T \in P_1 \text{ for all } T \in \mathcal{T}_h\}$$

and

$$Q_h = \{q \in C(\overline{\Omega}) \cap L_0^2(\Omega) : \ q|_T \in P_1 \text{ for all } T \in \mathcal{T}_h\}.$$

Observe that, contrary to the Q_1-P_0 element or the P_1-P_0 element, here the pressure is approximated by a continuous function.

The spaces V_h and Q_h fulfill the corresponding approximation properties on regular meshes. However, the element is not stable.

In order to stabilize the element the space V_h is enlarged.

Consider an arbitrary triangle $T \in \mathcal{T}_h$ with vertices x_i, $i = 1, 2, 3$. Each point $x \in T$ can uniquely be represented in the form

$$x = \sum_{i=1}^{3} \lambda_i \, x_i$$

with

$$\lambda_i \geq 0, \quad \sum_{i=1}^{3} \lambda_i = 1.$$

The coefficients λ_i are called the barycentric coordinates of x. We introduce the following function:

$$b_T(x) = \lambda_1 \, \lambda_2 \, \lambda_3.$$

Obviously we have: $b_T \in P_3$. Because of the property $b_T(x) = 0$ for all $x \in \partial T$ the function b_T is called a bubble function.

The following extension of V_h is introduced:

$$\overline{V}_h = \{v \in C_0(\overline{\Omega}, \mathbb{R}^2) : \ v|_T = p_T + b_T \, \beta_T, \ p_T \in P_1, \ \beta_T \in \mathbb{R}^2 \text{ for all } T \in \mathcal{T}_h\}.$$

We have

Theorem 3.3. *For regular triangular subdivisions the spaces $\overline{V}_h$ and Q_h satisfy the discrete inf-sup condition with a constant independent of h.*

Sketch of the proof: We use Lemma 3.1 and construct a linear and bounded operator $\Pi_h \colon V \to \overline{V}_h$, such that

$$\begin{aligned} 0 &= \int_\Omega q_h \operatorname{div}(v_h - v)\ dx = -\int_\Omega (v_h - v) \cdot \operatorname{grad} q_h\ dx \\ &= -\sum_{T \in \mathcal{T}_h} \operatorname{grad} q_h \cdot \int_\Omega (v_h - v)\ dx \quad \text{for all } q_h \in Q_h \end{aligned}$$

with $v_h = \Pi_h v$. It suffices to satisfy the following condition:

$$\int_T v_h\ dx = \int_T v\ dx \quad \text{for all } T \in \mathcal{T}_h.$$

Let Δ_i be the union of all triangles from $\mathcal{T}_h$, which contain x_i as a vertex. Let $v_h \in \overline{V}_h$ be given by

$$v_h(x_i) = \frac{1}{\Delta_i} \int_{\Delta_i} v\ dx \quad \text{for } i = 1, 2, 3$$

and

$$\int_T v_h\ dx = \int_T v\ dx.$$

It is easy to show that v_h is well-defined and the operator Π_h is linear.

By a so-called scaling argument one shows the existence of an h-independent constant C with

$$\|v_h\|_1 \le C \|v\|_1.$$

□

The additional degrees of freedom b_T by adding the bubble functions β_T can be locally eliminated (static condensation):

With the ansatz

$$u_h = u_h^1 + u_h^b, \quad u_h^1 \in V_h,\ u_h^b = \sum_{T \in \mathcal{T}_h} b_T\, \beta_T \in (B_3)^2$$

where

$$B_3 = \operatorname{span}\{b_T : T \in \mathcal{T}_h\}$$

one obtains from

$$a(u_h, v_h) + b(v_h, p_h) = \langle F, v_h \rangle$$

for the test functions $b_T\, e_i$, $i = 1, 2$:

$$\sum_{j=1}^{2} a(b_T\, e_j, b_T\, e_i)\, \beta_{T,j} + a(u_h^1, b_T\, e_i) + b(b_T\, e_i, p_h) = \langle F, b_T\, e_i \rangle \quad \text{for } i = 1, 2.$$

So the values $\beta_{T,j}$ can be expressed in terms of the restriction of the unknowns u_h^1 and p_h on the triangle T.

Especially for the Stokes equation we have:

$$a(b_T\, e_j, b_T\, e_i) = \delta_T\, \delta_{ij} \quad \text{with} \quad \delta_T = \nu \int_T \| \operatorname{grad}(b_T)\|_{\ell_2}^2 \, dx$$

and

$$a(v_h, w_h) = a(w_h, v_h) = 0 \quad \text{for all } v_h \in V_h,\ w_h \in (B_3)^2.$$

Hence

$$\delta_T\, \beta_T - \int_T p_h \operatorname{grad} b_T \, dx = \int_T b_T\, f \, dx.$$

This implies

$$\beta_T = \frac{1}{\delta_T} \int_T (b_T\, f + p_h \operatorname{grad} b_T) \, dx = \frac{1}{\delta_T} \int_T b_T\, (f - \operatorname{grad} p_h) \, dx = \frac{\gamma_T}{\delta_T} (\overline{f}^T - \operatorname{grad} p_h)$$

with

$$\gamma_T = \int_T b_T \, dx, \quad \overline{f}^T = \frac{1}{\gamma_T} \int_T b_T\, f \, dx.$$

Using the ansatz for the second equation

$$-\int_\Omega q_h \operatorname{div} u_h \, dx = 0$$

we obtain

$$\begin{aligned}
0 &= -\int_\Omega q_h \operatorname{div} u_h^1 \, dx - \int_\Omega q_h \operatorname{div} u_h^b \, dx \\
&= -\int_\Omega q_h \operatorname{div} u_h^1 \, dx - \sum_{T \in \mathcal{T}_h} \int_T q_h \operatorname{div}(b_T\, \beta_T) \, dx \\
&= -\int_\Omega q_h \operatorname{div} u_h^1 \, dx + \sum_{T \in \mathcal{T}_h} \int_T b_T\, \beta_T \cdot \operatorname{grad} q_h \, dx \\
&= -\int_\Omega q_h \operatorname{div} u_h^1 \, dx + \sum_{T \in \mathcal{T}_h} \frac{\gamma_T}{\delta_T} (\overline{f}^T - \operatorname{grad} p_h) \cdot \operatorname{grad} q_h \int_T b_T \, dx \\
&= -\int_\Omega q_h \operatorname{div} u_h^1 \, dx + \sum_{T \in \mathcal{T}_h} \alpha(T) \int_T (\overline{f}^T - \operatorname{grad} p_h) \cdot \operatorname{grad} q_h \, dx
\end{aligned}$$

with

$$\alpha(T) = \frac{\gamma_T^2}{\delta_T\, |T|} = O(h_T^2).$$

Together with

$$a(u_h, v_h^1) + b(v_h^1, p_h) = a(u_h^1, v_h^1) + b(v_h^1, p_h) = \langle F, v_h^1 \rangle \quad \text{for all } v_h^1 \in V_h$$

one obtains a mixed variational problem in the original spaces, however, with a modified second equation:

$$\begin{aligned} a(u_h^1, v_h^1) + b(v_h^1, p_h) &= \langle F, v\rangle \qquad \text{for all } v_h^1 \in V_h, \\ b(u_h^1, q_h) - c_h(p_h, q_h) &= \langle G_h, q_h\rangle \quad \text{for all } q_h \in Q_h \end{aligned}$$

with

$$\begin{aligned} c_h(p_h, q_h) &= \sum_{T\in\mathcal{T}_h} \alpha(T) \int_T \operatorname{grad} p_h \cdot \operatorname{grad} q_h \; dx, \\ \langle G_h, q_h\rangle &= \sum_{T\in\mathcal{T}_h} \alpha(T) \int_T \overline{f}^T \cdot \operatorname{grad} q_h \; dx. \end{aligned}$$

So, adding the bubble functions corresponds to a modification of the second equation with a new mesh-dependent bilinear form c_h and a mesh-dependent linear functional G_h .

3.3.4 The Taylor-Hood Element

As an example of an element with higher accuracy we consider the Taylor-Hood element on a triangular subdivision $\mathcal{T}_h$ of $\Omega \subset \mathbb{R}^2$. Let

$$V_h = \{v \in C_0(\overline{\Omega}, \mathbb{R}^2) : \; v|_T \in P_2 \text{ for all } T \in \mathcal{T}_h\}$$

and

$$Q_h = \{q \in C(\overline{\Omega}) \cap L_0^2(\Omega) : \; q|_T \in P_1 \text{ for all } T \in \mathcal{T}_h\}.$$

Of course, we have

$$V_h \subset V = H_0^1(\Omega, \mathbb{R}^2) \quad \text{and} \quad Q_h \subset Q = L_0^2(\Omega).$$

These spaces satisfy the following approximation properties:

1. For $u \in H_0^1(\Omega, \mathbb{R}^2)$ and $p \in L_0^2(\Omega)$ we have:

$$\lim_{h\to 0} \inf_{v_h\in V_h} \|u - v_h\|_1 = 0 \quad \text{and} \quad \lim_{h\to 0} \inf_{q_h\in Q_h} \|p - q_h\|_0 = 0.$$

2. Under the stronger assumptions $u \in H_0^1(\Omega, \mathbb{R}^2) \cap H^3(\Omega.\mathbb{R}^2)$ and $p \in L_0^2(\Omega) \cap H^2(\Omega)$ there exists a constant C with

$$\inf_{v_h\in V_h} \|u - v_h\|_1 \le C\, h^2 \, \|u\|_3 \quad \text{and} \quad \inf_{q_h\in Q_h} \|p - q_h\|_0 \le C\, h^2 \, \|p\|_2.$$

The inf-sup condition can be shown by the so-called Verfürth trick, which consists of two steps:

Lemma 3.2. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with Lipschitz-continuous boundary. Let $V_h \subset V = H_0^1(\Omega, \mathbb{R}^d)$ and $Q_h \subset Q = L_0^2(\Omega) \cap H^1(\Omega)$ be closed subspaces. Assume that there exists a linear operator $R_h : V \longrightarrow V_h$ and a constant c independent of h such that*

$$\left(\sum_{T \in \mathcal{T}_h} h_T^{-2} \left\| v - R_h v \right\|_{0,T}^2 \right)^{1/2} \leq c_0 \left\| v \right\|_1 \quad \textit{and} \quad \left\| R_h v \right\|_1 \leq c_1 \left\| v \right\|_1.$$

Then there exist two positive constants c_2 and c_3 such that

$$\sup_{v_h \in V_h} \frac{\int_\Omega q_h \operatorname{div} v_h \, dx}{\left\| v_h \right\|_1} \geq c_2 \left\| q_h \right\|_0 - c_3 \left(\sum_{T \in \mathcal{T}_h} h_T^2 \left\| \operatorname{grad} q_h \right\|_{0,T}^2 \right)^{1/2}.$$

Proof. The inf-sup condition holds:

$$\inf_{q \in L_0^2(\Omega)} \sup_{v \in V} \frac{\int_\Omega q \operatorname{div} v \, dx}{\left\| v \right\|_1 \left\| q \right\|_0} \geq \beta_1 > 0.$$

Therefore, for each $q_h \in Q_h$ there exists a $\bar{v} \in V$ with

$$\frac{\int_\Omega q_h \operatorname{div} \bar{v} \, dx}{\left\| \bar{v} \right\|_1} \geq \frac{\beta_1}{2} \left\| q_h \right\|_0.$$

This implies

$$\begin{aligned}
\sup_{v_h \in V_h} \frac{\int_\Omega q_h \operatorname{div} v_h \, dx}{\left\| v_h \right\|_1} &\geq \frac{\max(0, \int_\Omega q_h \operatorname{div} R_h \bar{v} \, dx)}{\left\| R_h \bar{v} \right\|_1} \\
&\geq \frac{\int_\Omega q_h \operatorname{div} R_h \bar{v} \, dx}{c_1 \left\| \bar{v} \right\|_1} \\
&= \frac{\int_\Omega q_h \operatorname{div} \bar{v} \, dx}{c_1 \left\| \bar{v} \right\|_1} + \frac{\int_\Omega q_h \operatorname{div}(R_h \bar{v} - \bar{v}) \, dx}{c_1 \left\| \bar{v} \right\|_1} \\
&\geq \frac{\beta_1}{2c_1} \left\| q_h \right\|_0 - \frac{\int_\Omega \operatorname{grad} q_h \cdot (R_h \bar{v} - \bar{v}) \, dx}{c_1 \left\| \bar{v} \right\|_1}
\end{aligned}$$

Now we have

$$\begin{aligned}\int_\Omega \operatorname{grad} q_h \cdot (R_h\bar{v} - \bar{v})\ dx &= \sum_{T\in\mathcal{T}_h} \int_T \operatorname{grad} q_h \cdot (R_h\bar{v} - \bar{v})\ dx \\ &\le \sum_{T\in\mathcal{T}_h} h_T \,\|\operatorname{grad} q_h\|_{0,T}\ h_T^{-1}\,\|\bar{v} - R_h\bar{v}\|_{0,T} \\ &\le \left(\sum_{T\in\mathcal{T}_h} h_T^2\,\|\operatorname{grad} q_h\|_{0,T}^2\right)^{1/2} \left(\sum_{T\in\mathcal{T}_h} h_T^{-2}\,\|\bar{v} - R_h\bar{v}\|_{0,T}^2\right)^{1/2} \\ &\le c_0 \left(\sum_{T\in\mathcal{T}_h} h_T^2\,\|\operatorname{grad} q_h\|_{0,T}^2\right)^{1/2} \|\bar{v}\|_1\end{aligned}$$

Hence

$$\sup_{v_h\in V_h} \frac{\int_\Omega q_h\ \operatorname{div} v_h\ dx}{\|v_h\|_1} \ \ge\ \frac{\beta_1}{2c_1}\|q_h\|_0 - \frac{c_0}{c_1}\left(\sum_{T\in\mathcal{T}_h} h_T^2\,\|\operatorname{grad} q_h\|_{0,T}^2\right)^{1/2}$$

□

The operator R_h is called a Clément operator.

The second step of Verfürth's trick is contained in the next lemma:

Lemma 3.3. *Let $\mathcal{T}_h$ be a regular triangular subdivision of Ω with the property that each element $T \in \mathcal{T}_h$ has at least two internal edges. Then there exists a positive constant c_4 with*

$$\sup_{v_h\in V_h} \frac{\int_\Omega q_h\ \operatorname{div} v_h\ dx}{\|v_h\|_1} \ge c_4\ \left(\sum_{T\in\mathcal{T}_h} h_T^2\,\|\operatorname{grad} q_h\|_{0,T}^2\right)^{1/2}$$

Proof. We have

$$\int_\Omega q_h\ \operatorname{div} v_h\ dx = -\int_\Omega \operatorname{grad} q_h \cdot v_h\ dx = -\sum_{T\in\mathcal{T}_h} \operatorname{grad} q_h \cdot \int_T v_h\ dx$$

Let $\mathcal{E}_h$ be the set of all internal edges of triangles in $\mathcal{T}_h$. To each edge $E \in \mathcal{E}_h$ a parallel unit vector t_E is assigned.

Let $q_h \in Q_h$ be arbitrary but fixed. Let $v_h \in V_h$ be that function which vanishes in all vertices of triangles in $\mathcal{T}_h$ and which fulfills:

$$v_h(m_E) = -h_E^2\,(\operatorname{grad} q_h \cdot t_E)\,t_E,$$

where m_E denotes the midpoint of the edge E.

Since $v_h \in P_2$ on T, it follows that

$$\int_T v_h \, dx = \frac{|T|}{3} \sum_{E \subset \partial T} v_h(m_E).$$

Hence

$$\begin{aligned}
\int_\Omega q_h \, \operatorname{div} v_h \, dx &= -\sum_{T \in \mathcal{T}_h} \operatorname{grad} q_h \cdot \int_T v_h \, dx = \sum_{T \in \mathcal{T}_h} \sum_{E \subset \partial T} \frac{1}{3} |T| h_E^2 \, (\operatorname{grad} q_h \cdot t_E)^2 \\
&\geq c \sum_{T \in \mathcal{T}_h} |T| h_E^2 \, \| \operatorname{grad} q_h \|_{\ell_2}^2 = c_1 \sum_{T \in \mathcal{T}_h} h_E^2 \, \| \operatorname{grad} q_h \|_{0,T}^2 \\
&\geq c_1 c_2 \sum_{T \in \mathcal{T}_h} h_T^2 \, \| \operatorname{grad} q_h \|_{0,T}^2
\end{aligned}$$

because

$$\sum_{E \subset \partial T} (t_E \cdot z)^2 \geq c_1 \, \|z\|_{\ell_2}^2 \quad \text{and} \quad \min_{E \subset T} h_E \geq c_2 \, h_T.$$

as a consequence of the regularity of the subdivision.

Furthermore, the regularity of the subdivision implies

$$\|v_h\|_{1,T}^2 \leq c \, h_T^{-2} \, |T| \sum_{E \subset \partial T} |v_h(m_E)|^2.$$

Therefore,

$$\begin{aligned}
\|v_h\|_1^2 &= \sum_{T \in \mathcal{T}_h} \|v_h\|_{1,T}^2 \leq c \sum_{T \in \mathcal{T}_h} h_T^{-2} \, |T| \sum_{E \subset \partial T} |v_h(m_E)|^2 \\
&= c \sum_{T \in \mathcal{T}_h} h_T^{-2} \, |T| \sum_{E \subset \partial T} h_E^4 \, (\operatorname{grad} q_h \cdot t_E)^2 \leq c' \sum_{T \in \mathcal{T}_h} h_T^2 \, |T| \, \| \operatorname{grad} q_h \|_{\ell_2}^2 \\
&= c' \sum_{T \in \mathcal{T}_h} h_T^2 \, \| \operatorname{grad} q_h \|_{0,T}^2
\end{aligned}$$

□

3.4 Mixed FEM for the Hellinger-Reissner Formulation

For simplicity only the case of pure Dirichlet boundary condition with $u_D = 0$ is considered:

Find $\sigma \in V$ and $u \in Q = L^2(\Omega, \mathbb{R}^3)$ such that

$$\begin{aligned}
a(\sigma, \tau) + b(\tau, u) &= 0 && \text{for all } \tau \in V, \\
b(\sigma, v) &= \langle G, v \rangle && \text{for all } v \in Q
\end{aligned}$$

with

$$a(\sigma,\tau)=\int_\Omega C^{-1}\sigma:\tau\ dx,\quad b(\tau,u)=\int_\Omega \operatorname{div}\tau\cdot u\ dx,\quad \langle G,v\rangle=-\int_\Omega f\cdot v\ dx$$

and

$$V=\{\tau\in H(\operatorname{div},\Omega,\mathbb{S})\mid \int_\Omega \operatorname{trace}\tau\ dx=0\}.$$

By Galerkin's principle appropriate subspaces $V_h\subset V$, $Q_h\subset Q$ are chosen and the approximate solution $(\sigma_h,u_h)\in V_h\times Q_h$ is given by the variational problem

$$\begin{array}{lll} a(\sigma_h,\tau_h)+b(\tau_h,u_h) & =0 & \text{for all } \tau_h\in V_h,\\ b(\sigma_h,v_h) & =\langle G,v_h\rangle & \text{for all } v_h\in Q_h. \end{array}$$

As discussed in the last chapter the coercivity of a on the kernel $\operatorname{Ker}B_h$ and the discrete inf-sup condition of b with a constant independent of h do not automatically follow from the corresponding conditions of the continuous problem.

This time the coercivity of a on V does not hold. Coercivity of a could only be shown on the set

$$W=\operatorname{Ker}B=\{\tau\in V\mid b(\tau,v)=0 \text{ for all } v\in Q\}=\{\tau\in V\mid \operatorname{div}\tau=0\}.$$

Only for the case that

$$W_h=\operatorname{Ker}B_h\subset W=\operatorname{Ker}B$$

with

$$W_h=\operatorname{Ker}B_h=\{\tau_h\in V_h\mid b(\tau_h,v_h)=0 \text{ for all } v_h\in Q_h\}$$

one immediately obtains the coercivity of a on $\operatorname{Ker}B_h$ with $\widetilde{\alpha}_1=\alpha_1$. All requirements formulated so far are not easy to fulfill for simple piecewise polynomial finite element spaces.

A mixed element for triangular subdivisions in $\mathbb{R}^2$, which satisfies all these requirements is defined as follows, see D. A. Arnold, R. Winther, 2001:

$$V_h=\{\tau\in V\colon \tau\big|_T\in P_3 \text{ and } \operatorname{div}\tau\big|_T\in P_1 \text{ for all } T\in\mathcal{T}_h\}$$

and

$$Q_h=\{v\in Q\colon v\big|_T\in P_1 \text{ for all } T\in\mathcal{T}_h\}$$

An element τ_h in V_h is given by

1. the values of τ_h at the vertices,
2. the values $\int_S \tau_h n\ ds$ and $\int_S \tau_h n\ s\ ds$ on each edge S and
3. the value $\int_T \tau_h\ dx$ on each triangle T.

This results in 24 degrees of freedom on each triangle.

Another possibility is to abandon the strong formulation of the symmetry of the stress tensor. So far, the starting point of the variational formulation was the following classical formulation of the linear elasticity problem:

$$\begin{aligned} C^{-1}\,\sigma - \varepsilon(u) &= 0 \quad && \text{in } \Omega, \\ \operatorname{div}\sigma &= -f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma. \end{aligned}$$

From the first equation the symmetry of σ immediately follows. Now we have

$$\varepsilon(u) = \nabla u - \omega(u)$$

with

$$\omega(u)_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} - \frac{\partial u_j}{\partial x_i}\right).$$

Observe that $\omega(u)$ is an anti-symmetric tensor. This motivates the following equivalent formulation of the first equation:

$$C^{-1}\,\sigma - \nabla u + \gamma = 0, \quad \sigma^T = \sigma.$$

Obviously, $\gamma = \omega(u)$ is the only anti-symmetric tensor that satisfies the first equation, which then is equivalent to the original first equation.

So the new starting point of a mixed variational formulation is the following system:

$$\begin{aligned} C^{-1}\,\sigma - \nabla u + \gamma &= 0 \quad && \text{in } \Omega, \\ \sigma^T - \sigma &= 0 && \text{in } \Omega, \\ \operatorname{div}\sigma &= -f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma. \end{aligned}$$

By multiplying the first equation component-wise by a test function τ, integrating over Ω and adding up, one obtains:

$$\int_\Omega C^{-1}\sigma : \tau \; dx - \int_\Omega \tau : \nabla u \; dx + \int_\Omega \tau : \gamma \; dx = 0.$$

By integration by parts it follows that:

$$\int_\Omega C^{-1}\sigma : \tau \; dx + \int_\Omega \operatorname{div}\tau \cdot u \; dx + \int_\Omega \tau : \gamma \; dx = 0.$$

By multiplying the second equation component-wise with an anti-symmetric tensor η, one obtains:

$$\int_\Omega \sigma^T : \eta \; dx - \int_\Omega \sigma : \eta \; dx = 0,$$

which is equivalent to

$$\int_\Omega \sigma : \eta \; dx = 0,$$

because η is anti-symmetric. Finally one obtains from the third equation

$$\int_\Omega \operatorname{div} \sigma \cdot v \; dx = -\int_\Omega f \cdot v \; dx.$$

By adding, one obtains

$$\int_\Omega \operatorname{div} \sigma \cdot v \; dx + \int_\Omega \sigma : \eta \; dx = -\int_\Omega f \cdot v \; dx.$$

Therefore, the following mixed variational problem results:

Find $\sigma \in V$ and $(u, \gamma) \in Q$ such that

$$\begin{array}{lll} a(\sigma,\tau) + b(\tau,(u,\gamma)) & = 0 & \text{for all } \tau \in V, \\ b(\sigma,(v,\eta)) & = \langle G, v\rangle & \text{for all } (v,\eta) \in Q \end{array}$$

with

$$a(\sigma,\tau) = \int_\Omega C^{-1}\sigma : \tau \; dx, \quad b(\tau,(u,\gamma)) = \int_\Omega \operatorname{div}\tau \cdot u \; dx + \int_\Omega \tau : \gamma \; dx,$$
$$\langle G, v\rangle = -\int_\Omega f \cdot v \; dx$$

and the spaces

$$V = H(\operatorname{div}, \Omega, \mathbb{R}^{3\times 3}), \quad Q = L^2(\Omega, \mathbb{R}^3) \times \{\gamma \in L^2(\Omega, \mathbb{R}^{3\times 3}) \,|\, \gamma + \gamma^T = 0\}.$$

The best-known element in $\mathbb{R}^2$ which is based on (the two-dimensional analogue of) this variational formulation is the PEERS-Element (plane elasticity element with reduced symmetry) for a triangular subdivision of Ω. It consists of the following components:

1. The Raviart-Thomas element of degree 0 (the RT_0 element) enlarged by functions, piecewise given by
$$c_T \operatorname{curl} b_T \quad \text{with } c_T \in \mathbb{R},$$
where $b_T = \lambda_1 \lambda_2 \lambda_3$ denotes the bubble-function on a triangle T, for the rows of σ_h;
2. Piecewise constant functions (the P_0 element) for u_h;
3. Continuous and piecewise linear functions (the P_1 element) for γ_h.

The Raviart-Thomas element of degree 0 (the RT_0 element) is an $H(\operatorname{div}, \Omega)$-conforming element, piecewise given by functions of the form

$$a_T + d_T\, x \quad \text{with } a_T \in \mathbb{R}^2,\ d_T \in \mathbb{R}.$$

The PEERS element is also suitable for almost incompressible materials.

Chapter 4

Solution of the Discretized Equations

The discussed mixed FEMs lead to linear systems of equations of the following form:

$$\begin{pmatrix} A & B^T \\ B & -C \end{pmatrix} \begin{pmatrix} u \\ p \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix}. \tag{4.1}$$

Throughout the chapter we will assume that A is symmetric and positive definite, that C is symmetric and positive semi-definite, and that the so-called (negative) Schur complement

$$S = C + BA^{-1}B^T$$

is non-singular.

Under these assumptions the matrix $\mathcal{K}$, given by

$$\mathcal{K} = \begin{pmatrix} A & B^T \\ B & -C \end{pmatrix},$$

is non-singular and allows a block LU factorization

$$\mathcal{K} = \begin{pmatrix} A & 0 \\ B & -S \end{pmatrix} \begin{pmatrix} I & A^{-1}B^T \\ 0 & I \end{pmatrix}. \tag{4.2}$$

The system (4.1) is equivalent to the following system:

$$\begin{aligned} Au + B^T p &= f, \\ Sp &= h \quad \text{with} \quad h = BA^{-1}f - g. \end{aligned}$$

4.1 The Uzawa Method and Variants

Let $p^{(0)}$ be a given initial guess for p. The classical Uzawa method is given by the following steps:

$$\begin{aligned} Au^{(k+1)} &= f - B^T p^{(k)}, \\ p^{(k+1)} &= p^{(k)} + \tau\,(Bu^{(k+1)} - Cp^{(k)} - g). \end{aligned}$$

with some positive parameter $\tau > 0$. It is clear that, in the case of convergence, the limit values solve the system (4.1).

If $u^{(k+1)}$ is eliminated, one obtains:

$$p^{(k+1)} = p^{(k)} + \tau\,(h - Sp^{(k)}),$$

which is the classical Richardson method applied to the system

$$Sp = h.$$

Since S is symmetric and positive definite, the convergence is guaranteed for sufficiently small parameters $\tau > 0$.

The convergence can be improved by using the preconditioned Richardson method with an appropriate preconditioner $\hat{S}$ for S. Then the iterative method reads

$$p^{(k+1)} = p^{(k)} + \hat{S}^{-1}(h - Sp^{(k)}).$$

In the original form we obtain the so-called preconditioned Uzawa method:

$$\begin{aligned} Au^{(k+1)} &= f - B^T p^{(k)}, \\ p^{(k+1)} &= p^{(k)} + \hat{S}^{-1}(Bu^{(k+1)} - Cp^{(k)} - g). \end{aligned}$$

An obvious disadvantage of the preconditioned Uzawa method is the necessity to compute $u^{(k+1)}$ as exact solution of the system

$$Au = b \quad \text{with} \quad b = f - B^T p^{(k)}.$$

If instead one step of some preconditioned Richardson method is used for determining $u^{(k+1)}$

$$u^{(k+1)} = u^{(k)} + \hat{A}^{-1}(b - Au^{(k)}),$$

where $\hat{A}$ is an appropriate preconditioner for A, then one obtains a so-called inexact preconditioned Uzawa method (also called preconditioned Arrow-Hurwicz method):

$$\begin{aligned} u^{(k+1)} &= u^{(k)} + \hat{A}^{-1}(f - Au^{(k)} - B^T p^{(k)}), \\ p^{(k+1)} &= p^{(k)} + \hat{S}^{-1}(Bu^{(k+1)} - Cp^{(k)} - g). \end{aligned}$$

Hence

$$\begin{aligned} \hat{A}(u^{(k+1)} - u^{(k)}) &= f - Au^{(k)} - B^T p^{(k)}, \\ B(u^{(k+1)} - u^{(k)}) - \hat{S}(p^{(k+1)} - p^{(k)}) &= g - Bu^{(k)} + Cp^{(k)}. \end{aligned}$$

That is

$$\hat{\mathcal{L}} \begin{pmatrix} u^{(k+1)} - u^{(k)} \\ p^{(k+1)} - p^{(k)} \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} - \mathcal{K} \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix}$$

or, equivalently,

$$\begin{pmatrix} u^{(k+1)} \\ p^{(k+1)} \end{pmatrix} = \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} + \hat{\mathcal{L}}^{-1} \left[\begin{pmatrix} f \\ g \end{pmatrix} - \mathcal{K} \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} \right]$$

with

$$\hat{\mathcal{L}} = \begin{pmatrix} \hat{A} & 0 \\ B & -\hat{S} \end{pmatrix}.$$

So the inexact preconditioned Uzawa method can be interpreted as preconditioned Richardson method for (4.1) with the block triangular preconditioner $\hat{\mathcal{L}}$.

Remark: With the setting

$$\hat{A} = A, \qquad \hat{S} = \frac{1}{\tau} I$$

one obtains the classical Uzawa method. With the setting

$$\hat{A} = \frac{1}{\sigma} I, \qquad \hat{S} = \frac{1}{\tau} I$$

one obtains the classical Arrow-Hurwicz method.

Observe that the preconditioner $\hat{\mathcal{L}}$ is formally obtained from (4.2) by replacing A and S by $\hat{A}$ and $\hat{S}$ in the first factor and ignoring the second factor.

If instead the second factor in (4.2) is treated analogously to the first factor, then the preconditioner $\hat{\mathcal{K}}$ is obtained:

$$\hat{\mathcal{K}} = \begin{pmatrix} \hat{A} & 0 \\ B & -\hat{S} \end{pmatrix} \begin{pmatrix} I & \hat{A}^{-1} B^T \\ 0 & I \end{pmatrix} = \begin{pmatrix} \hat{A} & B^T \\ B & B\hat{A}^{-1} B^T - \hat{S} \end{pmatrix}.$$

Observe that $\hat{\mathcal{K}}$ is a symmetric and indefinite block matrix.

One step of the preconditioned Richardson method

$$\begin{pmatrix} u^{(k+1)} \\ p^{(k+1)} \end{pmatrix} = \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} + \hat{\mathcal{K}}^{-1} \left[\begin{pmatrix} f \\ g \end{pmatrix} - \mathcal{K} \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} \right]$$

requires the solution of the system

$$\hat{\mathcal{K}} \begin{pmatrix} u^{(k+1)} - u^{(k)} \\ p^{(k+1)} - p^{(k)} \end{pmatrix} = \begin{pmatrix} f \\ g \end{pmatrix} - \mathcal{K} \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix}.$$

This can be done in three steps:

$$\begin{aligned} \hat{A} \, (\hat{u}^{(k+1)} - u^{(k)}) &= f - A u^{(k)} - B^T p^{(k)}, \\ \hat{S} \, (p^{(k+1)} - p^{(k)}) &= B \hat{u}^{(k+1)} - C p^{(k)} - g, \\ \hat{A} \, (u^{(k+1)} - u^{(k)}) &= f - A u^{(k)} - B^T p^{(k+1)}. \end{aligned}$$

Interpretation: From the first and the third equation one obtains:

$$u^{(k+1)} = \hat{u}^{(k+1)} - \hat{A}^{-1}B^T(p^{(k+1)} - p^{(k)}). \tag{4.3}$$

(4.3) is considered as ansatz for the next approximation $u^{(k+1)}$ which is required to solve the equation

$$Bu^{(k+1)} - Cp^{(k+1)} = g.$$

This leads to the equation:

$$H(p^{(k+1)} - p^k) = B\hat{u}_{k+1} - Cp^{(k)} - g$$

with

$$H = C + B\hat{A}^{-1}B^T.$$

H is called the so-called inexact Schur complement. If compared with the second equation, one could interpret $\hat{S}$ as preconditioner for H and the second equation is just one step of the corresponding preconditioned Richardson method applied to the equation

$$Hp' = c \quad \text{with} \quad c = B\hat{u}_{k+1} - Cp^{(k)} - g$$

with starting value 0 for computing $p' = p^{(k+1)} - p^k$.

Remark: For the case $C = 0$ and the choice $\hat{S} = H = B\hat{A}^{-1}B^T$, i.e.:

$$\mathcal{K} = \begin{pmatrix} A & B^T \\ B & 0 \end{pmatrix} \quad \text{and} \quad \hat{\mathcal{K}} = \begin{pmatrix} \hat{A} & B^T \\ B & 0 \end{pmatrix},$$

the preconditioned Richardson method

$$\begin{pmatrix} u^{(k+1)} \\ p^{(k+1)} \end{pmatrix} = \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} + \hat{\mathcal{K}}^{-1}\left[\begin{pmatrix} f \\ g \end{pmatrix} - \mathcal{K}\begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix}\right]$$

can also be written as a projection method:

$$u^{(k+1)} = P(u^{(k)} + \hat{A}^{-1}[f - Au^{(k)}]).$$

Here P is the $\hat{A}$-orthogonal projection on the linear manifold $V_g = \{v \in \mathbb{R}^n : Bv = g\}$, i.e.: $w = Pu \in V_g$ is the unique solution of the variational problem

$$(w, v)_{\hat{A}} = (u, v)_{\hat{A}} \quad \text{for all } v \in V_0 = \operatorname{Ker} B.$$

4.2 Preconditioner for the Schur Complement

An approximate solution of the mixed variational problem

$$\begin{array}{lcll} a(u,v)+b(v,p) & = & \langle F,v\rangle & \text{for all } v\in V \\ b(u,q) & = & \langle G,q\rangle & \text{for all } q\in Q \end{array}$$

is obtained by an appropriate choice of finite-dimensional subspaces

$$V_h \subset V, \quad Q_h \subset Q.$$

By Galerkin's principle the approximate solutions $u_h \in V_h$ and $p_h \in Q_h$ are the solutions of the discrete variational problem

$$\begin{array}{lcll} a(u_h,v_h)+b(v_h,p_h) & = & \langle F,v_h\rangle & \text{for all } v_h\in V_h \\ b(u_h,q_h) & = & \langle G,q_h\rangle & \text{for all } q_h\in Q_h. \end{array}$$

Let $\{\varphi_j\}$ be a basis for V_h and let $\{\psi_k\}$ be a basis for Q_h. Then the approximate solution can be represented in the following way:

$$u_h = \sum_j u_j\ \varphi_j, \quad p_h = \sum_k p_k\ \psi_k.$$

From the discrete variational problem one obtains the following linear system of equations

$$\begin{pmatrix} A_h & B_h^T \\ B_h & 0 \end{pmatrix} \begin{pmatrix} \underline{u}_h \\ \underline{p}_h \end{pmatrix} = \begin{pmatrix} \underline{f}_h \\ \underline{g}_h \end{pmatrix}$$

with

$$\begin{array}{lcl} A_h & = & (a(\varphi_j,\varphi_i))\,, \\ B_h & = & (b(\varphi_j,\psi_k))\,, \\ \underline{u}_h & = & (u_j)\,, \quad \underline{p}_h = (p_k)\,, \quad \underline{f}_h = \quad (\langle F,\varphi_i\rangle)\,, \quad \underline{g}_h = \quad (\langle G,\psi_k\rangle)\,. \end{array}$$

Assume that the following conditions are satisfied:

1. The bilinear form a is symmetric, coercive and bounded on V. Then $\|v\|_V = a(v,v)^{1/2}$ can be chosen as a norm in V.

2. The bilinear form b is bounded:

$$|b(v,q)| \le \beta_2\, \|v\|_V \|q\|_Q.$$

3. The discrete inf-sup condition is satisfied:

$$\inf_{0\neq q_h\in Q_h} \sup_{0\neq v_h\in V_h} \frac{b(v_h,q_h)}{\|v_h\|_V\ \|q_h\|_Q} \ge \widetilde{\beta}_1 > 0,$$

where $\widetilde{\beta}$ is independent of h.

Then we have:

$$\begin{aligned}(B_h A_h^{-1} B_h^T \ \underline{q}_h, \underline{q}_h)_{\ell_2} &= (A_h^{-1/2} B_h^T \ \underline{q}_h, A_h^{-1/2} B_h^T \ \underline{q}_h)_{\ell_2} \\ &= \sup_{\underline{w}_h \neq 0} \frac{(A_h^{-1/2} B_h^T \ \underline{q}_h, \underline{w}_h)_{\ell_2}^2}{(\underline{w}_h, \underline{w}_h)_{\ell_2}} = \sup_{\underline{v}_h \neq 0} \frac{(B_h \ \underline{v}_h, \underline{q}_h)_{\ell_2}^2}{(A_h \underline{v}_h, \underline{v}_h)_{\ell_2}} \\ &= \sup_{0 \neq v_h \in V_h} \frac{b(v_h, q_h)^2}{a(v_h, v_h)} = \sup_{0 \neq v_h \in V_h} \frac{b(v_h, q_h)^2}{\|v_h\|_V^2}.\end{aligned}$$

Therefore, the following estimates hold:

$$\widetilde{\beta}_1^2 \, \|q_h\|_Q^2 \leq (B_h A_h^{-1} B_h^T \ \underline{q}_h, \underline{q}_h)_{\ell_2} \leq \beta_2^2 \, \|q_h\|_Q^2.$$

Now

$$\|q_h\|_Q^2 = (M_h \underline{q}_h, \underline{q}_h)_{\ell_2} \quad \text{with} \quad M_h = \left((\psi^{(k)}, \psi^{(l)})_Q \right).$$

Hence

$$\widetilde{\beta}_1^2 \, M_h \leq S_h = B_h A_h^{-1} B_h^T \leq \beta_2^2 \, M_h.$$

The spectral constants $\widetilde{\beta}_1^2$ and β_2^2 are independent of h. Therefore, M_h is a spectrally equivalent preconditioner of the Schur complement $S_h = B_h A_h^{-1} B_h^T$. The corresponding preconditioned Uzawa method has the convergence rate

$$q = \frac{\kappa(M_h^{-1} S_h) - 1}{\kappa(M_h^{-1} S_h) + 1} \leq \frac{(\beta_2 / \widetilde{\beta}_1)^2 - 1}{(\beta_2 / \widetilde{\beta}_1)^2 + 1} < 1,$$

which is independent of h.

Application to the Stokes problem

For $\|v\|_V = |v|_1$ and for pure Dirichlet boundary conditions we have

$$b(v, q) = -\int_\Omega q \, \mathrm{div} \, v \ dx \leq \|q\|_0 \| \, \mathrm{div} \, v\|_0 \leq \|q\|_0 |v|_1.$$

Hence: $\beta_2 = 1$.

The matrix M_h is the mass matrix, which is spectrally equivalent to $h^d \, I$ for regular meshes:

$$c_1 h^d \, I \leq M_h \leq c_2 h^d \, I.$$

Therefore, the Uzawa method (without preconditioner) for $\tau = O(h^d)$ converges optimally (h-independent convergence rate). If instead of Richardson's method the gradient method is applied to $Sp = h$ it suffices to set the parameter $\tau = 1$.

The method can be additionally accelerated by the CG method.

4.3 Convergence Analysis for Inexact Uzawa Methods

We have

$$\begin{aligned}\begin{pmatrix} u^{(k+1)} \\ p^{(k+1)} \end{pmatrix} &= \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} + \hat{\mathcal{L}}^{-1} \left[\begin{pmatrix} f \\ g \end{pmatrix} - \mathcal{K} \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} \right] \\ &= \mathcal{M} \begin{pmatrix} u^{(k)} \\ p^{(k)} \end{pmatrix} + \hat{\mathcal{L}}^{-1} \begin{pmatrix} f \\ g \end{pmatrix}\end{aligned}$$

with the iteration matrix

$$\mathcal{M} = I - \hat{\mathcal{L}}^{-1}\mathcal{K}.$$

Therefore, it follows for the error

$$e^{(j)} = \begin{pmatrix} u^{(j)} - u^* \\ p^{(j)} - p^* \end{pmatrix}$$

that

$$e^{(k+1)} = \mathcal{M} e^{(k)}.$$

Now

$$\begin{aligned}\mathcal{M} &= \hat{\mathcal{L}}^{-1}(\hat{\mathcal{L}} - \mathcal{K}) \\ &= \begin{pmatrix} \hat{A}^{-1} & 0 \\ \hat{S}^{-1}B\hat{A}^{-1} & -\hat{S}^{-1} \end{pmatrix} \begin{pmatrix} \hat{A} - A & -B^T \\ 0 & -\hat{S} + C \end{pmatrix} \\ &= \begin{pmatrix} \hat{A}^{-1}(\hat{A} - A) & -\hat{A}^{-1}B^T \\ \hat{S}^{-1}B\hat{A}^{-1}(\hat{A} - A) & I - \hat{S}^{-1}[C + B\hat{A}^{-1}B^T] \end{pmatrix} \\ &= \begin{pmatrix} -\hat{A}^{-1} & -\hat{A}^{-1}B^T\hat{S}^{-1} \\ -\hat{S}^{-1}B\hat{A}^{-1} & \hat{S}^{-1} - \hat{S}^{-1}[C + B\hat{A}^{-1}B^T]\hat{S}^{-1} \end{pmatrix} \begin{pmatrix} A - \hat{A} & 0 \\ 0 & \hat{S} \end{pmatrix} \\ &= \mathcal{N}\mathcal{Q}.\end{aligned}$$

This factorization of the iteration matrix into symmetric factors is the key for the convergence analysis, see [11] for typical results.

Here we concentrate on one special case: If

$$\hat{A} < A$$

then $\mathcal{Q}$ is symmetric and positive definite and, therefore, defines a new scalar product:

$$\left(\begin{pmatrix} u \\ p \end{pmatrix}, \begin{pmatrix} v \\ q \end{pmatrix} \right)_{\mathcal{Q}} = \left(\mathcal{Q} \begin{pmatrix} u \\ p \end{pmatrix}, \begin{pmatrix} v \\ q \end{pmatrix} \right)_{\ell_2} = ([A - \hat{A}]u, v)_{\ell_2} + (\hat{S}p, q)_{\ell_2}.$$

The iteration matrix $\mathcal{M}$ is symmetric with respect to this scalar product:

$$(\mathcal{M}x, y)_{\mathcal{Q}} = (\mathcal{Q}\mathcal{N}\mathcal{Q}x, y)_{\ell_2} = (\mathcal{Q}\mathcal{N}\mathcal{Q}y, x)_{\ell_2} = (\mathcal{M}y, x)_{\mathcal{Q}}.$$

Therefore, $\mathcal{M}$ and $\hat{\mathcal{L}}^{-1}\mathcal{K} = I - \mathcal{M}$ have only real eigenvalues. In particular the following convergence property can be shown:

Theorem 4.1. *If*

$$\hat{A} < A \leq \overline{\alpha}\,\hat{A} \quad \textit{and} \quad \underline{\sigma}\,\hat{S} \leq S \leq \overline{\sigma}\,\hat{S}$$

then the matrix $\hat{\mathcal{L}}^{-1}\mathcal{K}$ *is symmetric with respect to the scalar product* $(x,y)_{\mathcal{Q}}$ *and we have*

$$\sigma\left(\hat{\mathcal{L}}^{-1}\mathcal{K}\right) \subset [\underline{\lambda}, \overline{\lambda}] \subset (0, \infty)$$

with

$$\underline{\lambda} = \frac{1}{2}\left[\overline{\alpha}(1+\underline{\sigma}) - \sqrt{\overline{\alpha}^2(1+\underline{\sigma})^2 - 4\overline{\alpha}\,\underline{\sigma}}\,\right] \qquad \overline{\lambda} = \frac{1}{2}\left[\overline{\alpha}(1+\overline{\sigma}) + \sqrt{\overline{\alpha}^2(1+\overline{\sigma})^2 - 4\overline{\alpha}\,\overline{\sigma}}\,\right]$$

Proof. Let $\varphi(\lambda)$ be the negative Schur complement of $\mathcal{K} - \lambda\,\hat{\mathcal{L}}$ or, in short:

$$\varphi(\lambda) = -\,\mathrm{Schur}(\mathcal{K} - \lambda\,\hat{\mathcal{L}}).$$

Then we have:

$$\begin{aligned}\varphi(\lambda) &= -\,\mathrm{Schur}\begin{pmatrix} A - \lambda\,\hat{A} & B \\ (1-\lambda)\,B & -C + \lambda\,\hat{S}\end{pmatrix} \\ &= C - \lambda\,\hat{S} + (1-\lambda)\,B(A - \lambda\,\hat{A})^{-1}B^T.\end{aligned}$$

It is immediately clear that

$$\varphi(0) = S > 0$$

and

$$\varphi(\lambda) > 0 \quad \text{for } \lambda \leq 0.$$

The first block-diagonal block $A - \lambda\,\hat{A}$ of the block matrix $\mathcal{K} - \lambda\,\hat{\mathcal{L}}$ is symmetric and positive definite for $\lambda \leq 0$.

So $A - \lambda\,\hat{A}$ and $\varphi(\lambda)$ are non-singular for $\lambda \leq 0$. This implies that the matrix $\mathcal{K} - \lambda\,\hat{\mathcal{L}}$ is non-singular in this case and, therefore, all eigenvalues of $\hat{\mathcal{L}}^{-1}\mathcal{K}$ are positive.

For $0 < \lambda \leq 1$ it follows that

$$0 < A - \lambda\,\hat{A} \leq \left(1 - \frac{\lambda}{\overline{\alpha}}\right) A$$

and, therefore,

$$\varphi(\lambda) \geq C - \lambda\,\hat{S} + \frac{1-\lambda}{1-\frac{\lambda}{\overline{\alpha}}}\,BA^{-1}B^T \geq \left[\frac{1-\lambda}{1-\frac{\lambda}{\overline{\alpha}}} - \frac{\lambda}{\underline{\sigma}}\right] S = \underline{\theta}(\lambda)\,S$$

with

$$\underline{\theta}(\lambda) = \frac{1-\lambda}{1-\frac{\lambda}{\overline{\alpha}}} - \frac{\lambda}{\underline{\sigma}}.$$

The smallest root of $\underline{\theta}(\lambda)$ is the smallest root $\underline{\lambda} < 1$ of the quadratic equation

$$\lambda^2 - \overline{\alpha}(1+\underline{\sigma})\lambda + \overline{\alpha}\,\underline{\sigma} = 0.$$

For $\overline{\alpha} < \lambda < \infty$ it follows

$$A - \lambda \hat{A} \leq \left(1 - \frac{\lambda}{\overline{\alpha}}\right) A < 0$$

and, therefore,

$$\varphi(\lambda) \leq C - \lambda \hat{S} + \frac{1-\lambda}{1-\frac{\lambda}{\overline{\alpha}}} B A^{-1} B^T \leq \left[\frac{1-\lambda}{1-\frac{\lambda}{\overline{\alpha}}} - \frac{\lambda}{\overline{\sigma}}\right] S = \overline{\theta}(\lambda)\, S$$

with

$$\overline{\theta}(\lambda) = \frac{1-\lambda}{1-\frac{\lambda}{\overline{\alpha}}} - \frac{\lambda}{\overline{\sigma}}.$$

The largest root of $\overline{\theta}(\lambda)$ is the largest root $\overline{\lambda}$ of the quadratic equation

$$\lambda^2 - \overline{\alpha}(1+\overline{\sigma})\lambda + \overline{\alpha}\,\overline{\sigma} = 0.$$

These estimates show that $A - \lambda \hat{A}$ and $\varphi(\lambda)$ are non-singular for $\lambda < \underline{\lambda}$ and for $\lambda > \overline{\lambda}$. This implies that $\mathcal{K} - \lambda \hat{\mathcal{L}}$ is non-singular. Therefore, λ is not an eigenvalue of $\hat{\mathcal{L}}^{-1}\mathcal{K}$. □

As a simple consequence the following sufficient condition for convergence is obtained:

$$\overline{\lambda} < 2, \quad \text{i.e.:} \quad \overline{\alpha}\,(2 + \overline{\sigma}) < 4.$$

In any case the problem is symmetric and positive definite with respect to the scalar product $(x, y)_{\mathcal{Q}}$ and, therefore, the CG method can be applied.

Remark: The statements of the last theorem date back to the work in [2], see also [11]. Similar results can be shown for the symmetric variant, see [11].

Bibliography

[1] Dietrich Braess. *Finite Elemente. Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*. Berlin: Springer, 2003.

[2] James H. Bramble and Joseph E. Pasciak. A preconditioning technique for indefinite systems resulting from mixed approximations of elliptic problems. *Math. Comput.*, 50(181):1–17, 1988.

[3] Franco Brezzi and Michel Fortin. *Mixed and hybrid finite element methods.* New York etc.: Springer-Verlag, 1991.

[4] G. Duvaut and J.L. Lions. *Inequalities in mechanics and physics*. Berlin-Heidelberg-New York: Springer-Verlag, 1976.

[5] Miloslav Feistauer. *Mathematical methods in fluid dynamics.* London: Longman Scientific & Technical. New York: Wiley, 1993.

[6] Vivette Girault and Pierre-Arnaud Raviart. *Finite element methods for Navier-Stokes equations. Theory and algorithms*. Berlin etc.: Springer-Verlag, 1986.

[7] U. Langer. Numerische Festkörpermechanik, 1997. Vorlesungsskriptum.

[8] Jerrold E. Marsden and Thomas J.R. Hughes. *Mathematical foundations of elasticity.* New York: Dover Publications, Inc., 1983.

[9] Jindřich Nečas. *Les méthodes directes en théorie des équations elliptiques*. Academia, Praha, and Masson et Cie, Editeurs, Paris, 1967.

[10] J.A. Nitsche. On Korn's second inequality. *RAIRO, Anal. Numér.*, 15:237–248, 1981.

[11] Walter Zulehner. Analysis of iterative methods for saddle point problems: A unified approach. *Math. Comput.*, 71(238):479–505, 2002.