A POSTERIORI ANALYSIS FOR PDE's.

S. Repin,

V.A. Steklov Institute of Mathematics in St.-Petersburg

- Mathematical models of viscous fluids
- Stokes problem
- Inf–Sup condition
- Existence of a saddle–point
- Estimates of the distance to the set of solenoidal fields
- A posteriori estimates for the Stokes problem
- Estimates for almost incompressible fluids
- Generalizations
- Nonlinear models of viscous fluids
- Antiplane flow of Bingham fluid
- Functional a posteriori estimates for generalized Newtonian fluids

9 A POSTERIORI ESTIMATES FOR NONLINEAR VARIATIONAL PROBLEMS

- Abstract variational problem
- Dual and bidual functionals
- Compound functionals
- Uniformly convex functionals
- General form of the functional a posteriori estimate
- Example. Diffusion problem with Robin boundary conditions

Preface

This is the electronic version of the lecture course

Mat-5.210 Special Course in Computational Mechanics Autumn 2006:
"A POSTERIORI ANALYSIS FOR PDE's"
http://math.tkk.fi/teaching/lme/

that was prepared for students and PhD students of the Department of Mathematics of Helsinki University of Technology and read in 2006.
In general, it is aimed to give an answer to the question "How to verify the accuracy of approximate solutions of partial differential equations computed by various numerical methods ?" A posteriori error estimates present a tool able to give an answer to the above question. Nowadays a posteriori estimates form a basis of many powerful numerical techniques and are widely used in computational technologies. Therefore, the purpose of the course is *to discuss the main lines in a posteriori analysis and explain their mathematical foundations.*

First, the "classical" a posteriori error estimation methods developed for finite element approximations are considered. However, the major part of the course is devoted to a new *functional approach* to a posteriori error estimation developed in the last decade. Basic ideas of this approach are first explained on the paradigm of a simple elliptic problem. Further exposition contains applications to particular classes of problems: diffusion, linear elasticity, Stokes, biharmonic, variational inequalities, etc.

The material is based on earlier lectures on a posteriori estimates and adaptive methods that has been read by the author at the University of Houston, USA (2002); University of Jyväskylä, Finland (2003), Radon Institute of Computational and Applied Mathematics (RICAM) in Linz (2005), and at St.-Petersburg Polytechnical University (2000-2003). However, in general, the course is new. A special attention is paid on a posteriori error estimation methods for two important classes of problems that are now in the focus of numerous researches in numerical analysis, namely to *mixed FE approximations* and *approximations in the theory of viscous fluids*.

Full list of references is given at the end of the text, but certain key publications are also cited in the respective places related to the topic discussed.

I wish to express my gratitude to Helsinki University of Technology for the invitation and hospitality. I am grateful to Prof. R. Stenberg for his kind support and interesting discussions that helped to create the course in the present form and also to Mr. A. Niemi and Mrs. Tuula–Donskoi for their help during my visit.

Sergey Repin Espoo, November 2006

Lecture 1

The goal of the lecture is to
provide a background information,
shortly discuss the a priori error estimation methods
and to give a concise overview of
first a posteriori error estimation methods

Lecture plan

- Error estimation problem in computer simulation;
- A priori approach to the error analysis for PDE's;
- A posteriori methods developed in 1900–1975:
  - Heuristic Runge's rule;
  - Prager and Synge estimate;
  - Estimate of Mikhlin;
  - Estimates of Ostrowski for contractive mappings;
  - Estimates based on monotonicity (Collatz);

Let us begin with a "philosophic" question:

WHAT THE NUMBERS COMPUTED INDEED MEAN?

To convince yourself that the question stated is worth thinking out, please make

Task 1. "Baby" coupled problem.

$$\mathbf{z'' - 9z' - 10z = 0}, \quad \mathbf{z = z(x)}, \quad \mathbf{x \in [0, 8]},$$
$$\mathbf{z(0) = 1}, \qquad \mathbf{z'(0) = a_{N-1} - a_N},$$

where $\mathbf{a}$ is a solution of the system of the dimensionality $\mathbf{N}$

$$\mathbf{Ba = f}, \quad \mathbf{b_{ij}} = \frac{\mathbf{2S_i^2 S_j^2}}{\pi} \int_{\mathbf{0}}^{\pi} \left(\sin(\mathbf{i}\xi)\sin(\mathbf{j}\xi) + \sin(\mathbf{i + j^2})\xi\right) \mathrm{d}\xi,$$

$$\mathbf{i, j = 1, 2, ...N}, \qquad \mathbf{f_i = (i + 1)^4 i}, \qquad \mathbf{S_i} = \sum_{\mathbf{k=0}}^{+\infty} \left(\frac{\mathbf{i}}{\mathbf{i + 1}}\right)^{\mathbf{k}}.$$

The task

For $\mathbf{N = 10, 50, 100, 200}$ find $\mathbf{z(8)}$ analytically and compare with numerical results obtained by computing the sums numerically, finding definite integrals with help of quadratures formulas, solving the system of linear simultaneous equations by a numerical method, and integrating the differential equation by a certain (e.g., Euler) method.

I. In the vast majority of cases, exact solutions of differential equations are unknown. We have no other way to use differential equations in the mathematical modeling other than compute approximate solutions and analyze computer simulation results.

II. Approximate solutions contain errors of various nature.

From **I** and **II**, it follows that

III. Verification of the accuracy of approximate solutions a KEY QUESTION.

Errors in mathematical modeling

- ε_1 – **error of a mathematical model used**
- ε_2 – **approximation error arising when a differential model is replaced by a discrete one;**
- ε_3 – **numerical errors arising when solving a discrete problem.**

MODELING ERROR

Let **U** be a physical value that characterizes some process and **u** be a respective value obtained from the mathematical model. Then the quantity

$$\varepsilon_1 = |\mathbf{U} - \mathbf{u}|$$

is an **error of the mathematical model**.

Mathematical model always presents an "abridged" version of a physical object.

Therefore, $\varepsilon_1 > 0$.

TYPICAL SOURCES OF MODELING ERRORS

(a) "Second order" phenomena are neglected in a mathematical model.

(b) Problem data are defined with an uncertainty.

(c) Dimension reduction is used to simplify a model.

APPROXIMATION ERROR

Let $\mathbf{u_h}$ be a solution on a mesh of the size $\mathbf{h}$. Then, $\mathbf{u_h}$ encompasses the **approximation error**

$$\varepsilon_2 = |\mathbf{u} - \mathbf{u_h}|.$$

Classical error control theory is mainly focused on approximation errors.

NUMERICAL ERRORS

Finite–dimensional problems are also solved approximately, so that instead of u_h we obtain u_h^ε. The quantity

$$\varepsilon_3 = |u_h - u_h^\varepsilon|$$

shows an error of the numerical algorithm performed with a concrete computer. This error includes

- **roundoff errors,**
- **errors arising in iteration processes and in numerical integration,**
- **errors caused by possible defects in computer codes.**

Roundoff errors

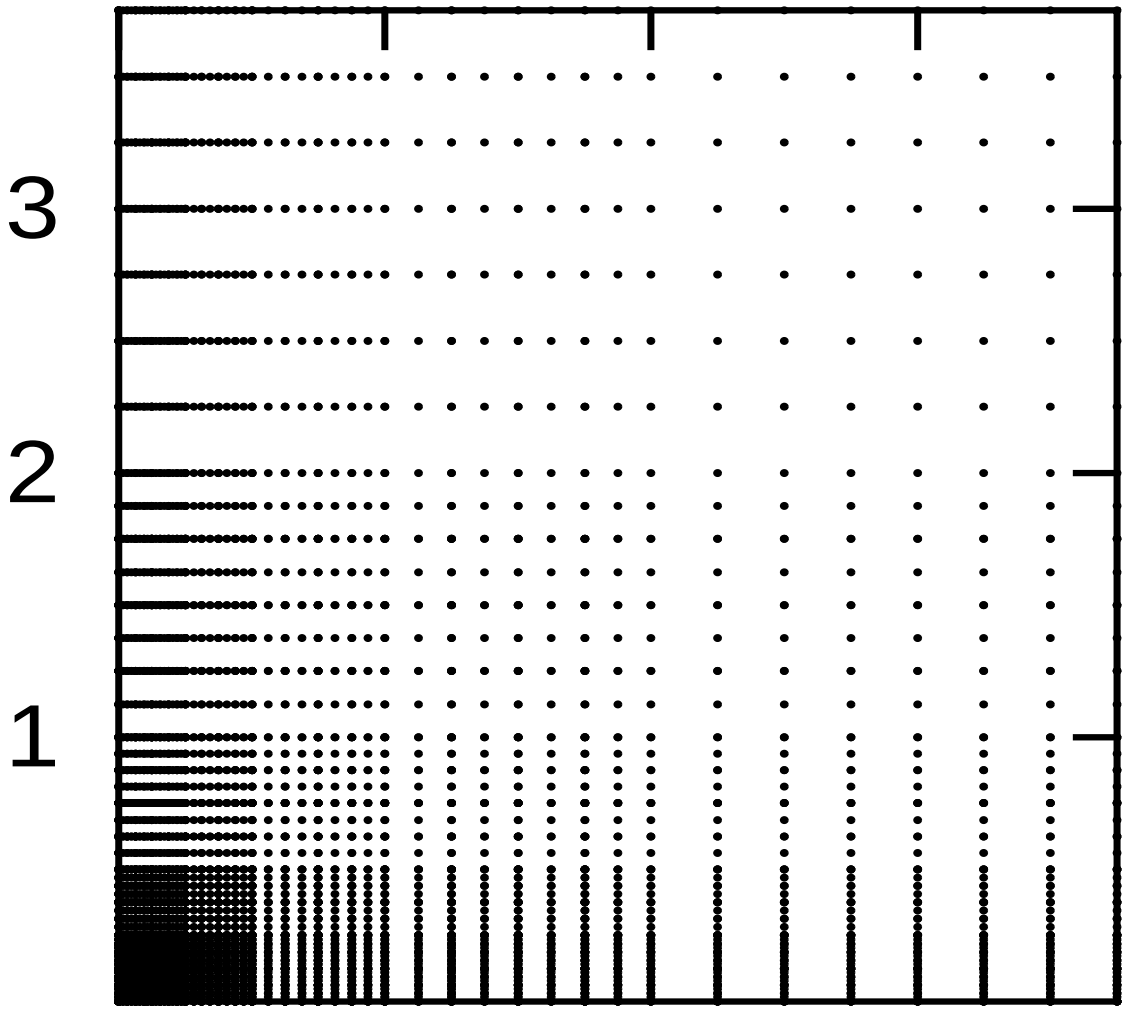
Numbers in a computer are presented in a floating point format:

$$\mathbf{x} = \pm \left(\frac{\mathbf{i_1}}{\mathbf{q}} + \frac{\mathbf{i_2}}{\mathbf{q^2}} + ... + \frac{\mathbf{i_k}}{\mathbf{q^k}} \right) \mathbf{q}^{\ell}, \quad \mathbf{i_s} < \mathbf{q}.$$

These numbers form the set $\mathbf{R_{q\ell k}} \subset \mathbb{R}$.
q is the **base** of the representation,
$\ell \in [\ell_1, \ell_2]$ is the **power**.

$\mathbf{R_{q\ell k}}$ is not closed with respect to the operations $+, -, *$!

The set $R_{q\ell k} \times R_{q\ell k}$



Example

$$\mathbf{k = 3}, \qquad \mathbf{a} = \left(\frac{\mathbf{1}}{\mathbf{2}} + \mathbf{0} + \mathbf{0}\right) * \mathbf{2^5}, \qquad \mathbf{b} = \left(\frac{\mathbf{1}}{\mathbf{2}} + \mathbf{0} + \mathbf{0}\right) * \mathbf{2^1}$$

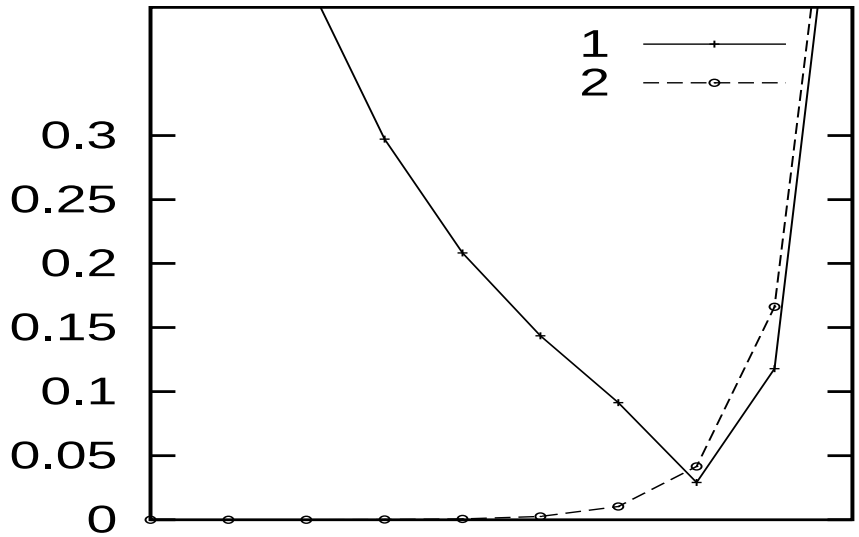
$$\mathbf{b} \Rightarrow \left(\mathbf{0} + \frac{\mathbf{1}}{\mathbf{2}} + \mathbf{0}\right) * \mathbf{2^2} \Rightarrow \left(\mathbf{0} + \mathbf{0} + \frac{\mathbf{1}}{\mathbf{2}}\right) * \mathbf{2^3} \Rightarrow (\mathbf{0} + \mathbf{0} + \mathbf{0}) * \mathbf{2^4}$$

$$\mathbf{a} + \mathbf{b} = \mathbf{a}!!!$$

Definition. The smallest floating point number which being added to 1 gives a quantity other than 1 is called **the machine accuracy**.

Numerical integration

$$\int_b^a f(x)dx \cong \sum_{i=1}^{n} c_i f(x_i)h = \sum_{i=1}^{n/2} c_i \overset{\sim 1}{f(x_i)}h + c_{n/2+1} \overset{\sim \delta}{f(x_{n/2+1})}h + \ldots$$



Errors in computer simulation

$\mathbf{U}$ **Physical object/process**

$\Downarrow$

$\varepsilon_1 \longrightarrow$ **Error of a model**

$\Downarrow$

$\mathbf{u}$ **Differential model** $\mathbf{Au = f}$

$\Downarrow$

$\varepsilon_2 \longrightarrow$ **Approximation error**

$\Downarrow$

$\mathbf{u_h}$ **Discrete model** $\mathbf{A^h u_h = f_h}$

$\Downarrow$

$\varepsilon_3 \longrightarrow$ **Computational error**

$\Downarrow$

$\mathbf{u_h^\varepsilon}$ **Numerical solution** $\mathbf{A^h u_h^\varepsilon = f_h} + \epsilon.$

Two principal relations

I. Computations on the basis of a reliable (certified) model. Here ε_1 is assumed to be small and u_h^ε gives a desired information on U.

$$\|\mathbf{U} - \mathbf{u}_\mathbf{h}^\varepsilon\| \leq \quad \varepsilon_1 + \boxed{\varepsilon_2 + \varepsilon_3}. \tag{1}$$

II. Verification of a mathematical model. Here physical data U and numerical data u_h^ε are compared to judge on the quality of a mathematical model

$$\|\varepsilon_1\| \leq \|\mathbf{U} - \mathbf{u}_\mathbf{h}^\varepsilon\| + \boxed{\varepsilon_2 + \varepsilon_3}. \tag{2}$$

Thus, two major problems of mathematical modeling, namely,

- reliable computer simulation,
- verification of mathematical models by comparing physical and mathematical experiments,

require efficient methods able to provide
COMPUTABLE AND REALISTIC

estimates of $\boxed{\varepsilon_2 + \varepsilon_3}$.

What is u and what is $\|\cdot\|$?

If we start a more precise investigation, then it is necessary to answer the question

What is a solution to a boundary–value problem?

Example.

$$\frac{\partial^2 \mathbf{u}}{\partial \mathbf{x}_1^2} + \frac{\partial^2 \mathbf{u}}{\partial \mathbf{x}_1^2} + \mathbf{f} = \mathbf{0}, \qquad \mathbf{u} = \mathbf{u_0} \text{ on } \partial\Omega \quad \boxed{\exists \mathbf{u}?}.$$

It is not a trivial question, so that about one hundred years passed before mathematicians have found an appropriate concept for PDE's.

Without proper understanding of a mathematical model no real modeling can be performed. Indeed,

If we are not sure that a solution u exists then what we try to approximate numerically?

If we do not know to which class of functions u belongs to, then we cannot properly define the measure for the accuracy of computed approximations.

Thus, we need to recall a
CONCISE MATHEMATICAL BACKGROUND

Vectors and tensors

$\mathbb{R}^n$ contains real n–vectors. $\mathbb{M}^{n\times m}$ contains $n \times m$ matrices and $\mathbb{M}^{n\times n}_s$ contains $n \times n$ symmetric matrices (tensors) with real entries.

$$\mathbf{a}\cdot\mathbf{b} = \sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{n}} \mathbf{a_i b_i} \in \mathbb{R}, \quad \mathbf{a},\mathbf{b}\in\mathbb{R}^{\mathbf{n}} \qquad \textbf{(scalar product of vectors)},$$

$$\mathbf{a}\otimes\mathbf{b} = \{\mathbf{a_i b_j}\} \in \mathbb{M}^{\mathbf{n}\times\mathbf{n}} \qquad \textbf{(tensor product of vectors)},$$

$$\boldsymbol{\sigma} : \boldsymbol{\varepsilon} = \sum_{\mathbf{i},\mathbf{j}=\mathbf{1}}^{\mathbf{n}} \boldsymbol{\sigma}_{\mathbf{ij}}\boldsymbol{\varepsilon}_{\mathbf{ij}} \in \mathbb{R}, \quad \boldsymbol{\sigma},\boldsymbol{\varepsilon}\in\mathbb{M}^{\mathbf{n}\times\mathbf{n}} \qquad \textbf{(scalar product of tensors)}.$$

$$|\mathbf{a}| := \sqrt{\mathbf{a}\cdot\mathbf{a}}, \qquad |\boldsymbol{\sigma}| := \sqrt{\boldsymbol{\sigma} : \boldsymbol{\sigma}},$$

Unit matrix is denoted by $\mathbb{I}$. If $\boldsymbol{\tau} \in \mathbb{M}^{n\times n}$, then $\boldsymbol{\tau}^D = \boldsymbol{\tau} - \frac{1}{n}\mathbb{I}$ is the deviator of $\boldsymbol{\tau}$.

Spaces of functions

Let Ω be an open bounded domain in $\mathbb{R}^n$ with Lipschitz continuous boundary.
$\mathbf{C^k(\Omega)}$ – k times continuously differentiable functions.
$\mathbf{C_0^k(\Omega)}$ – k times continuously differentiable functions vanishing at the boundary $\partial\Omega$.
$\mathbf{C_0^\infty(\Omega)}$ – k smooth functions with compact supports in Ω.
$\mathbf{L^p(\Omega)}$ – summable functions with finite norm

$$\|\mathbf{g}\|_{\mathbf{p},\Omega} = \|\mathbf{g}\|_{\mathbf{p}} = \left(\int_{\Omega} |\mathbf{g}|^{\mathbf{p}} \right)^{1/\mathbf{p}} .$$

For $\mathbf{L^2(\Omega)}$ the norm is denoted by $\|\cdot\|$.

If g is a vector (tensor)– valued function, then the respective spaces are denoted by
$\mathbf{C^k(\Omega, \mathbb{R}^n)}$ ($\mathbf{C^k(\Omega, \mathbb{M}^{n\times n})}$),
$\mathbf{L^p(\Omega, \mathbb{R}^n)}$ ($\mathbf{L^p(\Omega, \mathbb{M}^{n\times n})}$)
with similar norms.

We say that $\mathbf{g}$ is **locally integrable** in Ω and write $\mathbf{f} \in \mathbf{L^{1,loc}(\Omega)}$, if $\mathbf{g} \in \mathbf{L^1}(\omega)$ for any $\omega \subset\subset \Omega$. Similarly, one can define the space $\mathbf{L^{p,loc}(\Omega)}$ that consists of functions locally integrable with degree $\mathbf{p \geq 1}$.

Generalized derivatives

Let $\mathbf{f}, \mathbf{g} \in \mathbf{L}^{1,\mathbf{loc}}(\mathbf{\Omega})$ and

$$\int_{\Omega} \mathbf{g}\varphi \, \mathbf{dx} = -\int_{\Omega} \mathbf{f} \frac{\partial \varphi}{\partial \mathbf{x_i}} \, \mathbf{dx}, \quad \forall \varphi \in \mathring{\mathbf{C}}^{\mathbf{1}}(\mathbf{\Omega}).$$

Then $\mathbf{g}$ is called a **generalized derivative** (in the sense of Sobolev) of $\mathbf{f}$ with respect to x_i and we write

$$\boxed{\mathbf{g} = \frac{\partial \mathbf{f}}{\partial \mathbf{x_i}}.}$$

Higher order generalized derivatives

If $\mathbf{f}, \mathbf{g} \in \mathbf{L}^{1,\mathrm{loc}}(\Omega)$ and

$$\int_\Omega \mathbf{g}\varphi \, \mathbf{dx} = \int_\Omega \mathbf{f} \frac{\partial^2 \varphi}{\partial \mathbf{x_i} \partial \mathbf{x_j}} \, \mathbf{dx}, \quad \forall \varphi \in \mathring{\mathbf{C}}^2(\Omega),$$

then $\mathbf{g}$ is a generalized derivative of $\mathbf{f}$ with respect to x_i and x_j. For generalized derivatives we keep the classical notation and write

$$\mathbf{g} = \partial^2 \mathbf{f} / \partial \mathbf{x_i} \partial \mathbf{x_j} = \mathbf{f}_{,\mathbf{ij}}.$$

If $\mathbf{f}$ is differentiable in the classical sense, then its generalized derivatives coincide with the classical ones !

To extend this definition further, we use the multi-index notation and write $\mathbf{D^{\alpha} f}$ in place of $\partial^{\mathbf{k}}\mathbf{f}/\partial\mathbf{x}_1^{\alpha_1}\partial\mathbf{x}_2^{\alpha_2}\dots\partial\mathbf{x}_n^{\alpha_n}$.

Definition

Let $\mathbf{f}, \mathbf{g} \in \mathbf{L}^{1,\mathbf{loc}}(\Omega)$ and

$$\int_{\Omega} \mathbf{g}\varphi \, \mathbf{dx} = (-1)^{|\alpha|} \int_{\Omega} \mathbf{f}\, \mathbf{D}^{\alpha}\varphi \, \mathbf{dx}, \quad \forall \varphi \in \overset{\circ}{\mathbf{C}}{}^{\mathbf{k}}(\Omega).$$

Then, $\mathbf{g}$ is called a **generalized derivative** of $\mathbf{f}$ of degree

$$|\alpha| := \alpha_1 + \alpha_2 + \dots + \alpha_n$$

and we write

$$\boxed{\mathbf{g} = \mathbf{D}^{\alpha}\mathbf{f}}.$$

Sobolev spaces

S. L. Sobolev. *Some Applications of Functional Analysis in Mathematical Physics*, Izdt. Leningrad. Gos. Univ., Leningrad, 1955, English version: Translation of Mathematical Monographs, Volume 90, American Mathematical Society, Providence, RI, 1991.

Definition

The spaces of functions that have integrable generalized derivatives up to a certain order are called Sobolev spaces. $\mathbf{f} \in \mathbf{W^{1,p}(\Omega)}$ if $\mathbf{f} \in \mathbf{L^p}$ and all the generalized derivatives of $\mathbf{f}$ of the first order belong to L^p, i.e.,

$$\mathbf{f_{,i}} = \frac{\partial \mathbf{f}}{\partial \mathbf{x_i}} \in \mathbf{L^p(\Omega)}.$$

The norm in $\mathbf{W^{1,p}}$ is defined as follows:

$$\|\mathbf{f}\|_{\mathbf{1,p,\Omega}} := \left(\int_{\Omega} (|\,\mathbf{f}\,|^{\mathbf{p}} + \sum_{\mathbf{i=1}}^{\mathbf{n}} |\,\mathbf{f_{,i}}\,|^{\mathbf{p}}) \mathbf{dx} \right)^{\mathbf{1/p}}.$$

All the other Sobolev spaces are defined quite similarly: $\mathbf{f} \in \mathbf{W^{k,p}(\Omega)}$ if all generalized derivatives up to the order k are integrable with power p and the quantity

$$\|\mathbf{f}\|_{\mathbf{k,p,\Omega}} := \left(\int_{\mathbf{\Omega}} \sum_{|\,\boldsymbol{\alpha}\,| \leq \mathbf{k}} |\, \mathbf{D^{\alpha} f}\,|^{\mathbf{p}}\, \mathbf{dx} \right)^{\mathbf{1/p}}$$

is finite. For the Sobolev spaces $\mathbf{W^{k,2}(\Omega)}$ we also use a simplified notation $\mathbf{H^k(\Omega)}$.

Sobolev spaces of vector- and tensor-valued functions are introduced by obvious extensions of the above definitions. We denote them by $\mathbf{W^{k,p}(\Omega, \mathbb{R}^n)}$ and $\mathbf{W^{k,p}(\Omega, \mathbb{M}^{n \times n})}$, respectively.

Embedding Theorems

Relationships between the Sobolev spaces and $\mathbf{L^p(\Omega)}$ and $\mathbf{C^k(\Omega)}$ are given by **Embedding Theorems**.

If $p, q \geq 1$, $\ell > 0$ and $\ell + \frac{n}{q} \geq \frac{n}{p}$, then $\mathbf{W^{\ell,p}(\Omega)}$ is continuously embedded in $\mathbf{L^q(\Omega)}$. Moreover, if $\ell + \frac{n}{q} > \frac{n}{p}$, then the embedding operator is compact.

If $\ell - k > \frac{n}{p}$, then $\mathbf{W^{\ell,p}(\Omega)}$ is compactly embedded in $\mathbf{C^k(\overline{\Omega})}$.

Traces

The functions in Sobolev spaces have counterparts on $\partial\Omega$ called **traces**. Thus, there exist some bounded operators mapping the functions defined in Ω to functions defined on the boundary, e.g.,

$$\gamma : \mathbf{H^1}(\Omega) \rightarrow \mathbf{L^2}(\partial\Omega)$$

is called the **trace operator** if it satisfies the following conditions:

$$\gamma \mathbf{v} = \mathbf{v}\,|_{\partial\Omega}, \qquad \forall \mathbf{v} \in \mathbf{C^1}(\Omega),$$
$$\|\gamma \mathbf{v}\|_{\mathbf{2},\partial\Omega} \leq \mathbf{c}\|\mathbf{v}\|_{\mathbf{1},\mathbf{2},\Omega},$$

where $\mathbf{c}$ is a positive constant independent of $\mathbf{v}$. From these relations, we observe that such a trace is a natural generalization of the trace defined for a continuous function.

It was established that $\gamma \mathbf{v}$ forms a subset of $\mathbf{L^2}(\partial\mathbf{\Omega})$, which is the space $\mathbf{H^{1/2}}(\partial\mathbf{\Omega})$. The functions from other Sobolev spaces also are known to have traces in Sobolev spaces with fractional indices.

Henceforth, we understand the boundary values of functions in the sense of traces, so that

$$\mathbf{u} = \boldsymbol{\psi} \quad \text{on } \partial\mathbf{\Omega}$$

means that the trace $\gamma\mathbf{u}$ of a function $\mathbf{u}$ defined in Ω coincides with a given function $\boldsymbol{\psi}$ defined on $\partial\Omega$.

All the spaces of functions that have zero traces on the boundary are marked by the symbol $\circ$ (e.g., $\mathbf{\overset{\circ}{W}{}^{l,p}}(\mathbf{\Omega})$ and $\mathbf{\overset{\circ}{H}{}^{1}}(\mathbf{\Omega})$).

Inequalities

1. Friederichs-Steklov inequality.

$$\|\mathbf{w}\| \leq \mathbf{C_\Omega}\|\nabla \mathbf{w}\|, \quad \forall \mathbf{w} \in \mathbf{\mathring{H}^1(\Omega)}, \tag{3}$$

2. Poincaré inequality.

$$\|\mathbf{w}\| \leq \mathbf{\widetilde{C}_\Omega}\|\nabla \mathbf{w}\|, \quad \forall \mathbf{w} \in \mathbf{\widetilde{H}^1(\Omega)}, \tag{4}$$

where $\widetilde{H}^1(\Omega)$ is a subset of H^1 of functions with zero mean.

3. Korn's inequality.

$$\int_{\Omega} \left(|\mathbf{v}|^2 + |\varepsilon(\mathbf{v})|^2 \right) \mathbf{dx} \geq \mu_\Omega \|\mathbf{v}\|^2_{1,2,\Omega}, \ \forall \mathbf{v} \in \mathbf{H^1(\Omega, \mathbb{R}^n)}, \tag{5}$$

Generalized solutions

The concept of generalized solutions to PDE's came from
Petrov-Bubnov-Galerkin method.
B. G. Galerkin. Beams and plates. Series in some questions of elastic equilibrium of beams and plates (approximate translation of the title from Russian). *Vestnik Ingenerov, St.-Peterburg*, 19(1915), 897-908.

$$\int_{\Omega} (\Delta u + f) w \, dx = 0 \qquad \forall w$$

Integration by parts leads to the so–called **generalized formulation** of the problem: find $u \in \overset{\circ}{H}{}^1(\Omega) + u_0$ such that

$$\boxed{\int_{\Omega} \nabla u \cdot \nabla w \, dx = \int_{\Omega} f w \, dx \qquad \forall w \in \overset{\circ}{H}{}^1(\Omega)}$$

This idea admits wide extensions to many differential equations, see e.g., O. A. Ladyzhenskaya, *The boundary value problems of mathematical physics*. Springer-Verlag, New York, 1985

Definition

A symmetric form $\mathbf{B} : \mathbf{V} \times \mathbf{V} \rightarrow \mathbf{R}$, where V is a Hilbert space, called $V-$ *elliptic* if $\exists c_1 > 0, c_2 > 0$ such that

$$\mathbf{B}(\mathbf{u}, \mathbf{u}) \geq \mathbf{c_1} \|\mathbf{u}\|^2, \quad \forall \mathbf{u} \in \mathbf{V}$$

$$| \mathbf{B}(\mathbf{u}, \mathbf{v}) | \leq \mathbf{c_2} \|\mathbf{u}\| \|\mathbf{v}\|, \quad \forall \mathbf{u}, \mathbf{v} \in \mathbf{V}$$

General formulation for linear PDE's is: for a certain linear continuous functional $\mathbf{f}$ (from the space $\mathbf{V}^*$ topologically dual to $\mathbf{V}$) find $\mathbf{u}$ such that

$$\boxed{\mathbf{B}(\mathbf{u}, \mathbf{w}) = < \mathbf{f}, \mathbf{w} > \qquad \mathbf{w} \in \mathbf{V}.}$$

Existence of a solution

Usually, existence is proved by

Lax-Milgram Lemma For a bilinear form $\mathbf{B}$ there exists a linear bounded operator $\mathbf{A} \in \mathcal{L}(\mathcal{V}, \mathcal{V})$ such that

$$\mathbf{B}(\mathbf{u}, \mathbf{v}) = (\mathbf{Au}, \mathbf{v}), \quad \forall \mathbf{u}, \mathbf{v} \in \mathbf{V}$$

It has an inverse $\mathbf{A}^{-1} \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, such that $\|\mathbf{A}\| \leq \mathbf{c_2}$, $\|\mathbf{A}^{-1}\| \leq \frac{1}{\mathbf{c_1}}$.

We will follow another *modus operandi* !!!.

Variational approach

Lemma

If $\mathbf{J} : \mathbf{K} \to \mathbf{R}$ *is convex, continuous and coercive, i.e.,*

$$\mathbf{J}(\mathbf{w}) \to +\infty \qquad \text{as } \|\mathbf{w}\|_{\mathbf{V}} \to +\infty$$

and $\mathbf{K}$ *is a convex closed subset of a reflexive space* $\mathbf{V}$*, then the problem*

$$\inf_{\mathbf{w} \in \mathbf{K}} \mathbf{J}(\mathbf{w})$$

has a **minimizer** $\mathbf{u}$*. If* $\mathbf{J}$ *is strictly convex, then the minimizer is* **unique**.

See, e.g., I. Ekeland and R. Temam. *Convex analysis and variational problems*. North-Holland, Amsterdam, 1976.

Coercivity

Take $J(w) = \frac{1}{2}B(w,w) - < f, w >$ and let K be a certain subspace. Then

$$\frac{1}{2}B(w,w) \geq c_1 \|w\|_V^2, \qquad | < f, w > | \leq \|f\|_{V^*} \|w\|_V.$$

We see, that

$$J(w) \geq c_1 \|w\|_V^2 - \|f\|_{V^*} \|w\|_V \to +\infty \quad \text{as } \|w\|_V \to +\infty$$

Since J is strictly convex and continuous we conclude that a minimizer exists and unique.

Useful algebraic relation

First we present the algebraic identity

$$\frac{1}{2}\mathbf{B}(\mathbf{u}-\mathbf{v},\mathbf{u}-\mathbf{v}) = \frac{1}{2}\mathbf{B}(\mathbf{v},\mathbf{v}) - <\mathbf{f},\mathbf{v}> + \tag{6}$$

$$+ <\mathbf{f},\mathbf{u}> - \frac{1}{2}\mathbf{B}(\mathbf{u},\mathbf{u}) - \mathbf{B}(\mathbf{u},\mathbf{v}-\mathbf{u}) + <\mathbf{f},\mathbf{v}-\mathbf{u}> =$$

$$= \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) - \mathbf{B}(\mathbf{u},\mathbf{v}-\mathbf{u}) + <\mathbf{f},\mathbf{v}-\mathbf{u}>$$

From this identity we derive two important results:

- (a) Minimizer $\mathbf{u}$ satisfies $\mathbf{B}(\mathbf{u},\mathbf{w}) = <\mathbf{f},\mathbf{w}> \forall \mathbf{w}$;
- (b) Error is subject to the difference of functionals.

Let us show (a), i.e., that from (6) it follows the identity

$$\mathbf{B}(\mathbf{u}, \mathbf{v}-\mathbf{u}) =< \mathbf{f}, \mathbf{v}-\mathbf{u} > \qquad \forall \mathbf{v} \in \mathbf{K},$$

which is $\mathbf{B}(\mathbf{u}, \mathbf{w}) =< \mathbf{f}, \mathbf{w} >$ if set $\mathbf{w} = \mathbf{v} - \mathbf{u}$. Indeed, assume the opposite, i.e. $\exists \bar{\mathbf{v}} \in \mathbf{K}$ such that

$$\mathbf{B}(\mathbf{u}, \bar{\mathbf{v}}-\mathbf{u}) - < \mathbf{f}, \bar{\mathbf{v}}-\mathbf{u} >= \delta > \mathbf{0} \qquad (\bar{\mathbf{v}} \neq \mathbf{u}!)$$

Set $\widetilde{\mathbf{v}} := \mathbf{u} + \alpha(\bar{\mathbf{v}} - \mathbf{u})$, $\alpha \in \mathbb{R}$. Then $\widetilde{\mathbf{v}} - \mathbf{u} = \alpha(\bar{\mathbf{v}} - \mathbf{u})$ and

$$\frac{1}{2}\mathbf{B}(\mathbf{u}-\widetilde{\mathbf{v}}, \mathbf{u}-\widetilde{\mathbf{v}}) + \mathbf{B}(\mathbf{u}, \widetilde{\mathbf{v}}-\mathbf{u}) - < \mathbf{f}, \widetilde{\mathbf{v}}-\mathbf{u} >=$$

$$= \frac{\alpha^2}{2}\mathbf{B}(\bar{\mathbf{v}}-\mathbf{u}, \bar{\mathbf{v}}-\mathbf{u}) + \alpha\delta = \mathbf{J}(\widetilde{\mathbf{v}}) - \mathbf{J}(\mathbf{u}) \geq \mathbf{0}$$

However, for arbitrary α such an inequality cannot be true. Denote $\mathbf{a} = \mathbf{B}(\bar{\mathbf{v}} - \mathbf{u}, \bar{\mathbf{v}} - \mathbf{u})$. Then in the left–hand side we have a function $\mathbf{1}/\mathbf{2}\alpha^2\mathbf{a}^2 + \alpha\delta$, which always attains negative values for certain α. For example, set $\alpha = -\delta/\mathbf{a}^2$. Then, the left–hand side is equal to $-\frac{1}{2}\delta^2/\mathbf{a}^2 < \mathbf{0}$ and we arrive at a contradiction.

A priori approach to the error control problem

Error estimate

Now, we show (b). From

$$\frac{1}{2}\mathbf{B}(\mathbf{u}-\mathbf{v},\mathbf{u}-\mathbf{v}) = \\ = \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) - \mathbf{B}(\mathbf{u},\mathbf{v}-\mathbf{u}) + <\mathbf{f},\mathbf{v}-\mathbf{u}>$$

we obtain the error estimate:

$$\frac{1}{2}\mathbf{B}(\mathbf{u}-\mathbf{v},\mathbf{u}-\mathbf{v}) = \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}). \qquad (7)$$

See **S. G. Mikhlin.** *Variational methods in mathematical physics.* Pergamon, Oxford, 1964.
which immediately gives the **projection estimate**

Projection estimate

Let $\mathbf{u_h}$ be a minimizer of $\mathbf{J}$ on $\mathbf{K_h} \subset \mathbf{K}$. Then

$$\frac{1}{2}\mathbf{B}(\mathbf{u}-\mathbf{u_h},\mathbf{u}-\mathbf{u_h}) = \mathbf{J}(\mathbf{u_h}) - \mathbf{J}(\mathbf{u}) \leq \mathbf{J}(\mathbf{v_h}) - \mathbf{J}(\mathbf{u}) =$$

$$= \frac{1}{2}\mathbf{B}(\mathbf{u}-\mathbf{v_h},\mathbf{u}-\mathbf{v_h}) \quad \forall\, \mathbf{v_h} \in \mathbf{K_h}.$$

and we observe that

$$\boxed{\mathbf{B}(\mathbf{u}-\mathbf{u_h},\mathbf{u}-\mathbf{u_h}) = \inf_{\mathbf{v_h}\in\mathbf{K_h}} \mathbf{B}(\mathbf{u}-\mathbf{v_h},\mathbf{u}-\mathbf{v_h})} \tag{8}$$

Projection type estimates serve a basis for deriving **a priori convergence estimates**.

Interpolation in Sobolev spaces

Two key points: PROJECTION ESTIMATE and
INTERPOLATION IN SOBOLEV SPACES.
Interpolation theory investigates the difference between a function in a Sobolev space and its piecewise polynomial interpolant. Basic estimate on a simplex $\mathbf{T_h}$ is

$$|\mathbf{v} - \mathbf{\Pi_h v}|_{\mathbf{m,t,T_h}} \leq \mathbf{C(m,n,t)} \left(\frac{\mathbf{h}}{\boldsymbol{\rho}}\right)^{\mathbf{m}} \mathbf{h}^{\mathbf{2-m}} \|\mathbf{v}\|_{\mathbf{2,t,T_h}},$$

and on the whole domain

$$|\mathbf{v} - \mathbf{\Pi_h v}|_{\mathbf{m,t,\Omega_h}} \leq \mathbf{C h^{2-m}} \|\mathbf{v}\|_{\mathbf{2,t,\Omega_h}}.$$

Here $\mathbf{h}$ is a the element size and $\boldsymbol{\rho}$ is the inscribed ball diameter.

Asymptotic convergence estimates

Typical case is $\mathbf{m = 1}$ and $\mathbf{t = 2}$. Since

$$\mathbf{B(u - u_h, u - u_h) \le B(u - \Pi_h u, u - \Pi_h u) \le c_2 \|u - \Pi_h u\|^2}$$

for

$$\mathbf{B(w, w)} = \int_{\Omega} \nabla \mathbf{w} \cdot \nabla \mathbf{w} \, \mathbf{dx}$$

we find that

$$\|\nabla(\mathbf{u - u_h})\| \le \mathbf{Ch} |\mathbf{u}|_{2,2,\Omega}.$$

provided that

- **Exact solution is H^2 – regular;**
- **u_h is the Galerkin approximation;**
- **Elements do not "degenerate" in the refinement process.**

A priori convergence estimates cannot guarantee that the error **monotonically decreases** as $h \to 0$.
Besides, in practice we are interested in the error of a **concrete approximation on a particular mesh**. Asymptotic estimates could hardly be helpful in such a context because, in general, the constant **C** serves for the *whole class* of approximate solutions of a particular type. Typically it is either unknown or highly overestimated.

A priori convergence estimates have mainly a theoretical value: they show that an approximation method is correct "in principle.

For these reasons, a quite different approach to error control is rapidly developing. Nowadays it has already formed a new direction:

A Posteriori Error Control for PDE's .

A posteriori error estimation methods developed in 1900-1975

Runge's rule

At the end of 19th century a heuristic error control method was suggested by C. Runge who investigated numerical integration methods for ordinary differential equations.

Heuristic rule of C. Runge

If the difference between two approximate solutions computed on a coarse mesh $\mathcal{T}_h$ with mesh size $\mathbf{h}$ and refined mesh $\mathcal{T}_{h_{ref}}$ with mesh size $\mathbf{h_{ref}}$ (e.g., $\mathbf{h_{ref}} = \mathbf{h}/\mathbf{2}$) has become small, then both $\mathbf{u_{h_{ref}}}$ and $\mathbf{u_h}$ are probably close to the exact solution.

In other words, this rule can be formulated as follows:

$$\boxed{\text{If } \left[\mathbf{u_h} - \mathbf{u_{h_{ref}}}\right] \text{ is small then } \mathbf{u_{h_{ref}}} \text{ is close to } \mathbf{u}}$$

where $\left[\cdot\right]$ is a certain functional or mesh-dependent norm.

Also, the quantity $\left[u_h - u_{h_{ref}}\right]$ can be viewed (in terms of modern terminology) as a certain **a posteriori error indicator**.

Runge's heuristic rule is simple and was easily accepted by numerical analysts.

However, if we do not properly define the quantity $[\,\cdot\,]$, for which $[u_h - u_{h_{ref}}]$ is small, then the such a principle may be not true.

One can present numerous examples where **two subsequent elements of an approximation sequence are close to each other, but far from a certain joint limit**. For example, such cases often arise in the minimization (maximization) of functionals with "saturation" type behavior or with a "sharp–well" structure. Also, the rule may lead to a wrong presentation if, e.g., the refinement has not been properly done, so that new trial functions were added only in subdomains were an approximation is almost coincide with the true solution. Then two subsequent approximations may be very close, but at the same time not close to the exact solution.

Also, in practice, we need to now precisely what the word "close" means, i.e. we need to have a more concrete presentation on the error. For example, it would be useful to establish the following rule:

$$\textbf{If} \quad [u_h - u_{h_{ref}}] \leq \varepsilon \quad \textbf{then} \quad \|u_h - u\| \leq \delta(\varepsilon),$$

where the function $\delta(\varepsilon)$ is known and computable.

In subsequent lectures we will see that for a wide class of boundary–value problems it is indeed possible to derive such type generalizations of the Runge's rule.

Prager and Synge estimates

W. Prager and J. L. Synge. Approximation in elasticity based on the concept of function spaces, *Quart. Appl. Math.* 5(1947)

Prager and Synge derived an estimate on the basis of purely geometrical grounds. In modern terms, there result for the problem

$$\begin{aligned} \mathbf{\Delta u} + \mathbf{f} &= \mathbf{0}, \qquad \text{in } \mathbf{\Omega}, \\ \mathbf{u} &= \mathbf{0}, \qquad \text{on } \partial\mathbf{\Omega} \end{aligned}$$

reads as follows:

$$\|\nabla(\mathbf{u}-\mathbf{v})\|^2 + \|\nabla\mathbf{u}-\boldsymbol{\tau}\|^2 = \|\nabla\mathbf{v}-\boldsymbol{\tau}\|^2,$$

where $\boldsymbol{\tau}$ is a function satisfying the equation $\mathbf{div}\boldsymbol{\tau} + \mathbf{f} = \mathbf{0}$.
We can easily prove it by the **orthogonality relation**

$$\int_{\mathbf{\Omega}} \nabla(\mathbf{u}-\mathbf{v}) \cdot (\nabla\mathbf{u}-\boldsymbol{\tau})\, \mathbf{dx} = \mathbf{0} \qquad (\mathbf{div}(\nabla\mathbf{u}-\boldsymbol{\tau}) = \mathbf{0}\,!).$$

Estimate of Mikhlin

S. G. Mikhlin. *Variational methods in mathematical physics*. Pergamon, Oxford, 1964.

A similar estimate was derived by variational arguments (see Lecture 1). It is as follows:

$$\frac{1}{2}\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \le \mathbf{J}(\mathbf{v}) - \mathbf{inf}\mathbf{J},$$

where

$$\mathbf{J}(\mathbf{v}) := \frac{1}{2}\|\nabla \mathbf{v}\|^2 - (\mathbf{f},\mathbf{v}), \qquad \mathbf{inf}\mathbf{J} := \inf_{\mathbf{v}\in \overset{\circ}{\mathbf{H}}_1(\Omega)} \mathbf{J}(\mathbf{v}).$$

Dual problem

Since

$$\mathbf{inf J} = \sup_{\tau \in \mathbf{Q_f}} \left\{ -\frac{1}{2} \|\tau\|^2 \right\},$$

where

$$\mathbf{Q_f} := \left\{ \tau \in \mathbf{L_2}(\Omega, \mathbf{R^d}) \mid \int_\Omega \tau \cdot \nabla \mathbf{w} \, \mathbf{dx} = \int_\Omega \mathbf{f w} \, \mathbf{dx} \qquad \forall \mathbf{w} \in \mathring{\mathbf{H}}^{\mathbf{1}} \right\},$$

we find that

$$\frac{1}{2} \|\nabla(\mathbf{u} - \mathbf{v})\|^2 \leq \mathbf{J(v)} + \frac{1}{2} \|\tau\|^2, \qquad \forall \tau \in \mathbf{Q_f}.$$

Since

$$\begin{aligned} \mathbf{J}(\mathbf{v}) + \tfrac{1}{2}\|\boldsymbol{\tau}\|^2 &= \frac{1}{2}\|\nabla \mathbf{v}\|^2 - \int_\Omega \mathbf{f}\mathbf{v}\,\mathbf{dx} + \frac{1}{2}\|\boldsymbol{\tau}\|^2 = \\ &= \frac{1}{2}\|\nabla \mathbf{v}\|^2 - \int_\Omega \boldsymbol{\tau}\cdot\nabla \mathbf{v}\,\mathbf{dx} + \frac{1}{2}\|\boldsymbol{\tau}\|^2 = \\ &= \frac{1}{2}\|\nabla \mathbf{v} - \boldsymbol{\tau}\|^2 \end{aligned}$$

we arrive at the estimate

$$\boxed{\frac{1}{2}\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \le \frac{1}{2}\|\nabla \mathbf{v} - \boldsymbol{\tau}\|^2, \qquad \forall \boldsymbol{\tau} \in \mathbf{Q_f}.} \tag{9}$$

Difficulties

Estimates of Prager and Synge and of Mikhlin are valid for any $\mathbf{v} \in \overset{\circ}{\mathbf{H}}_1(\Omega)$, so that, formally, that they can be applied to any conforming approximation of the problem. However, from the practical viewpoint these estimates have an essential drawback:

> **they use a function τ in the set $\mathbf{Q_f}$ defined by the differential relation,**

which may be difficult to satisfy exactly. Probably by this reason further development of a posteriori error estimates for Finite Element Methods (especially in 80'-90') was mainly based on different grounds.

Fixed point theorem

Consider a Banach space $(\mathbf{X}, \mathbf{d})$ and a continuous operator

$$\mathfrak{T} : \mathbf{X} \to \mathbf{X}.$$

Definition

A point $\mathbf{x}_\odot$ is called a fixed point of $\mathfrak{T}$ if

$$\mathbf{x}_\odot = \mathfrak{T}\mathbf{x}_\odot \,. \tag{10}$$

Approximations of a fixed point are usually constructed by the iteration sequence

$$\mathbf{x_i} = \mathfrak{T}\mathbf{x_{i-1}} \qquad \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots \,. \tag{11}$$

Contractive mappings

Two basic tasks:

(a) find the conditions that guarantee convergence of $\mathbf{x_i}$ to $\mathbf{x_\odot}$,

(b) find computable estimates of the error $\mathbf{e_i} = \mathbf{d}(\mathbf{x_i}, \mathbf{x_\odot})$.

Definition

An operator $\mathfrak{T} : \mathbf{X} \to \mathbf{X}$ is called q-contractive on a set $\mathbf{S} \subset \mathbf{X}$ if there exists a positive real number q such that the inequality

$$\mathbf{d}(\mathfrak{T}\mathbf{x}, \mathfrak{T}\mathbf{y}) \leq \mathbf{q}\, \mathbf{d}(\mathbf{x}, \mathbf{y}) \tag{12}$$

holds for any elements $\mathbf{x}$ and $\mathbf{y}$ of the set $\mathbf{S}$.

Theorem (S. Banach)

Let $\mathfrak{T}$ be a $\mathbf{q}$-contractive mapping of a closed nonempty set $\mathbf{S} \subset \mathbf{X}$ to itself with $\mathbf{q} < \mathbf{1}$. Then, $\mathfrak{T}$ has a unique fixed point in $\mathbf{S}$ and the sequence $\mathbf{x_i}$ obtained by (11) converges to this point.

Proof. It is easy to see that

$$\mathbf{d}(\mathbf{x_{i+1}}, \mathbf{x_i}) = \mathbf{d}(\mathfrak{T}\mathbf{x_i}, \mathfrak{T}\mathbf{x_{i-1}}) \leq \mathbf{q}\mathbf{d}(\mathbf{x_i}, \mathbf{x_{i-1}}) \leq \ldots \leq \mathbf{q^i}\mathbf{d}(\mathbf{x_1}, \mathbf{x_0}).$$

Therefore, for any $\mathbf{m} > \mathbf{1}$ we have

$$\begin{aligned}\mathbf{d}(\mathbf{x_{i+m}}, \mathbf{x_i}) \leq \\ \leq \mathbf{d}(\mathbf{x_{i+m}}, \mathbf{x_{i+m-1}}) + \mathbf{d}(\mathbf{x_{i+m-1}}, \mathbf{x_{i+m-2}}) + \ldots + \mathbf{d}(\mathbf{x_{i+1}}, \mathbf{x_i}) \leq \\ \leq \mathbf{q^i}(\mathbf{q^{m-1}} + \mathbf{q^{m-2}} + \ldots + \mathbf{1})\mathbf{d}(\mathbf{x_1}, \mathbf{x_0}) \,. \quad (13)\end{aligned}$$

Since

$$\sum_{k=0}^{m-1} q^k \leq \frac{1}{1-q} \ ,$$

(13) implies the estimate

$$d(x_{i+m}, x_i) \ \leq \ \frac{q^i}{1-q} d(x_1, x_0). \qquad (14)$$

Let $i \to \infty$, then the right-hand side of (14) tends to zero, so that $\{x_i\}$ is a Cauchy sequence. It has a limit in $y \in X$.

Then, $\mathbf{d(x_i, y) \to 0}$ and

$$\mathbf{d(\mathfrak{T} x_i, \mathfrak{T} y) \leq q d(x_i, y) \to 0}$$

so that $\mathbf{d(\mathfrak{T} x_i, \mathfrak{T} y) \to 0}$ and $\mathbf{\mathfrak{T} x_i \to \mathfrak{T} y}$. Pass to the limit in (11) as $\mathbf{i \to +\infty}$. We observe that

$$\mathbf{\mathfrak{T} y = y}.$$

Hence, any limit of such a sequence is a fixed point.

It is easy to prove that a fixed point is unique.
Assume that there are two different fixed points $\mathbf{x}^1_\odot$ and $\mathbf{x}^2_\odot$, i.e.

$$\mathfrak{T}\mathbf{x}^k_\odot = \mathbf{x}^k_\odot, \qquad \mathbf{k} = \mathbf{1}, \mathbf{2}.$$

Therefore,

$$\mathbf{d}(\mathbf{x}^1_\odot, \mathbf{x}^2_\odot) = \mathbf{d}(\mathfrak{T}\mathbf{x}^1_\odot, \mathfrak{T}\mathbf{x}^2_\odot) \leq \mathbf{q}\mathbf{d}(\mathbf{x}^1_\odot, \mathbf{x}^2_\odot).$$

But $\mathbf{q} < \mathbf{1}$, and thus such an inequality cannot be true.

A priori convergence estimate

Let $\mathbf{e_j} = \mathbf{d}(\mathbf{x_j}, \mathbf{x_\odot})$ denote the error on the $\mathbf{j}$-th step. Then

$$\mathbf{e_j} = \mathbf{d}(\mathfrak{T}\mathbf{x_{j-1}}, \mathfrak{T}\mathbf{x_\odot}) \leq \mathbf{q}\mathbf{e_{j-1}} \leq \mathbf{q^j}\mathbf{e_0}.$$

and

$$\mathbf{e_j} \leq \mathbf{q^j}\mathbf{e_0}. \tag{15}$$

This estimate gives a certain presentation on that how the error decreases. However, this a priori upper bound may be rather coarse.

A posteriori estimates

The proposition below furnishes upper and lower estimates of $\mathbf{e_j}$, which are easy to compute provided, that the number $\mathbf{q}$ (or a good estimate of it) is known.

A. Ostrowski. Les estimations des erreurs a posteriori dans les procédés itératifs, *C.R. Acad.Sci. Paris Sér. A–B*, 275(1972), A275-A278.

Theorem (A. Ostrowski)

Let $\{x_j\}_{j=0}^{\infty}$ *be a sequence obtained by the iteration process*

$$x_i = \mathfrak{T} x_{i-1} \qquad i = 1, 2, \ldots$$

with a mapping $\mathfrak{T}$ *satisfying the condition* $\|\mathfrak{T}\| = q \leq 1$. *Then, for any* x_j, $j > 1$, *the following estimate holds:*

$$M^j_{\ominus} := \frac{1}{1+q} d(x_{j+1}, x_j) \leq e_j \leq M^j_{\oplus} := \frac{q}{1-q} d(x_j, x_{j-1}). \qquad (16)$$

Proof. The upper estimate in (16) follows from (14). Indeed, put $\mathbf{i = 1}$ in this relation. We have

$$\mathbf{d(x_{1+m}, x_1) \le \frac{q}{1-q} d(x_1, x_0)}\,.$$

Since $\mathbf{x_{1+m} \to x_\odot}$ as $\mathbf{m \to +\infty}$, we pass to the limit with respect to $\mathbf{m}$ and obtain

$$\mathbf{d(x_\odot, x_1) \le \frac{q}{1-q} d(x_1, x_0)}\,.$$

We may view $\mathbf{x_{j-1}}$ as the starting point of the sequence. Then, in the above relation $\mathbf{x_0 = x_{j-1}}$ and $\mathbf{x_1 = x_j}$ and we arrive at the following upper bound of the error:

$$\mathbf{d(x_\odot, x_j) \le \frac{q}{1-q} d(x_j, x_{j-1})}\,.$$

The lower bound of the error follows from the relation

$$\mathbf{d}(\mathbf{x_j}, \mathbf{x_{j-1}}) \leq \mathbf{d}(\mathbf{x_j}, \mathbf{x_\odot}) + \mathbf{d}(\mathbf{x_{j-1}}, \mathbf{x_\odot}) \leq (\mathbf{1}+\mathbf{q})\mathbf{d}(\mathbf{x_{j-1}}, \mathbf{x_\odot}),$$

which shows that

$$\mathbf{d}(\mathbf{x_{j-1}}, \mathbf{x_\odot}) \geq \frac{\mathbf{1}}{\mathbf{1}+\mathbf{q}} \mathbf{d}(\mathbf{x_j}, \mathbf{x_{j-1}}) \, .$$

Note that

$$\frac{\mathbf{M}^{\mathbf{j}}_{\oplus}}{\mathbf{M}^{\mathbf{j}}_{\ominus}} = \frac{\mathbf{q}(\mathbf{1}+\mathbf{q})}{\mathbf{1}-\mathbf{q}} \frac{\mathbf{d}(\mathbf{x_j}, \mathbf{x_{j-1}})}{\mathbf{d}(\mathbf{x_{j+1}}, \mathbf{x_j})} \geq \frac{\mathbf{1}+\mathbf{q}}{\mathbf{1}-\mathbf{q}} \, ,$$

we see that that the efficiency of the upper and lower bounds given by (16) deteriorates as $\mathbf{q} \to \mathbf{1}$.

Remark. If $\mathbf{X}$ is a normed space, then

$$\mathbf{d}(\mathbf{x_{j+1}}, \mathbf{x_j}) = \|\mathbf{R}(\mathbf{x_j})\| ,$$

where

$$\mathbf{R}(\mathbf{x_j}) := \mathfrak{T}\mathbf{x_j} - \mathbf{x_j}$$

is the residual of the basic equation (10). Thus, the upper and lower estimates of errors are expressed in terms of the **residuals of the respective iteration equation** computed for two neighbor steps:

$$\frac{\mathbf{1}}{\mathbf{1+q}} \|\mathbf{R}(\mathbf{x_j})\| \leq \mathbf{e_j} = \mathbf{d}(\mathbf{x_j}, \mathbf{x_\odot}) \leq \frac{\mathbf{q}}{\mathbf{1-q}} \|\mathbf{R}(\mathbf{x_{j-1}})\| .$$

Corollaries

In the iteration methods, it is often easier to analyze the operator

$$\mathfrak{T} = \mathbf{T}^{\mathbf{n}} := \underbrace{\mathbf{T}\mathbf{T}...\mathbf{T}}_{\mathbf{n} \text{ times}}$$

where $\mathbf{T}$ is a certain mapping.

Proposition (1)

Let $\mathbf{T} : \mathbf{S} \to \mathbf{S}$ *be a continuous mapping such that* $\mathfrak{T}$ *is a* $\mathbf{q}$*-contractive mapping with* $\mathbf{q} \in (\mathbf{0}, \mathbf{1})$*. Then, the equations*

$$\mathbf{x} = \mathbf{T}\mathbf{x} \qquad \text{and} \qquad \mathbf{x} = \mathfrak{T}\mathbf{x}$$

have one and the same fixed point, which is unique and can be found by the above described iteration procedure.

Proof. By the Banach Theorem, we observe that the operator $\mathfrak{T}$ has a unique fixed point $\boldsymbol{\xi}_{\odot}$.
Let us show that $\boldsymbol{\xi}_{\odot}$ is a fixed point of $\mathbf{T}$, we note that

$$\mathbf{T}\boldsymbol{\xi}_{\odot} = \mathbf{T}(\mathfrak{T}\boldsymbol{\xi}_{\odot}) = \mathbf{T}\,\mathfrak{T}^{2}\boldsymbol{\xi}_{\odot} = \ldots$$
$$= \mathbf{T}\,\mathfrak{T}^{\mathbf{i}}\boldsymbol{\xi}_{\odot} = \mathbf{T}^{(1+\mathbf{in})}\boldsymbol{\xi}_{\odot} = \mathbf{T}^{\mathbf{in}}\mathbf{T}\boldsymbol{\xi}_{\odot}. \quad (17)$$

Denote $\mathbf{x_0} = \mathbf{T}\boldsymbol{\xi}_{\odot}$. By (17) we conclude that for any $\mathbf{i}$

$$\mathbf{T}\boldsymbol{\xi}_{\odot} = \mathfrak{T}^{\mathbf{i}}\mathbf{x_0}. \quad (18)$$

Passing to the limit on the right-hand side in (18), we arrive at the relation $\mathbf{T}\boldsymbol{\xi}_{\odot} = \boldsymbol{\xi}_{\odot}$, which means that $\boldsymbol{\xi}_{\odot}$ is a fixed point of the operator $\mathbf{T}$.

Let $\widetilde{\mathbf{x}_\odot}$ be a fixed point of $\mathbf{T}$. Then,

$$\widetilde{\mathbf{x}_\odot} = \mathbf{T}^2\widetilde{\mathbf{x}_\odot} = .. = \mathbf{T}^n\widetilde{\mathbf{x}_\odot} = \mathfrak{T}\widetilde{\mathbf{x}_\odot}$$

and we observe that $\widetilde{\mathbf{x}_\odot}$ is a fixed point of $\mathbf{T}$. Since the saddle point of $\mathfrak{T}$ exists and is unique, we conclude that

$$\mathbf{x}_\odot = \widetilde{\mathbf{x}_\odot}.$$

Remark. This assertion may be practically useful if it is not possible to prove that $\mathbf{T}$ is $\mathbf{q}$–contractive, but this fact can be established for a certain power of $\mathbf{T}$.

Iteration methods for bounded linear operators

Consider a bounded linear operator $\mathcal{L} : \mathbf{X} \to \mathbf{X}$, where $\mathbf{X}$ is a Banach space. Given $\mathbf{b} \in \mathbf{X}$, the iteration process is defined by the relation

$$\mathbf{x_j} = \mathcal{L}\,\mathbf{x_{j-1}} + \mathbf{b}. \tag{19}$$

Let $\mathbf{x}_\odot$ be a fixed point of (19) and

$$\|\mathcal{L}\| = \mathbf{q} < \mathbf{1}.$$

By applying the Banach Theorem it is easy to show that

$$\{\mathbf{x_j}\} \to \mathbf{x}_\odot .$$

Indeed, let $\bar{\mathbf{x}}_\mathbf{j} = \mathbf{x_j} - \mathbf{x}_\odot$. Then

$$\bar{\mathbf{x}}_\mathbf{j} = \mathcal{L}\mathbf{x_{j-1}} + \mathbf{b} - \mathbf{x}_\odot = \mathcal{L}(\mathbf{x_{j-1}} - \mathbf{x}_\odot) = \mathcal{L}\bar{\mathbf{x}}_\mathbf{j-1} . \qquad (20)$$

Since

$$\mathbf{0_X} = \mathcal{L}\,\mathbf{0_X},$$

we note that the zero element $\mathbf{0_X}$ is a unique fixed point of the operator $\mathcal{L}$. By the Banach theorem $\bar{\mathbf{x}}_\mathbf{j} \to \mathbf{0_X}$ and, therefore, $\{\mathbf{x_j}\} \to \mathbf{x}_\odot$.

Therefore, we have an *a priori* estimate

$$\|\mathbf{x_j} - \mathbf{x_\odot}\|_{\mathbf{X}} = \|\bar{\mathbf{x}}_\mathbf{j} - \mathbf{0_X}\|_{\mathbf{X}} \leq$$

$$\leq \frac{\mathbf{q^j}}{\mathbf{1-q}} \|\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_0\|_{\mathbf{X}} = \frac{\mathbf{q^j}}{\mathbf{1-q}} \|\mathbf{R}(\mathbf{x_0})\|_{\mathbf{X}} \quad (21)$$

and the *a posteriori* one

$$\|\mathbf{x_j} - \mathbf{x_\odot}\|_{\mathbf{X}} \leq \frac{\mathbf{q}}{\mathbf{1-q}} \|\mathbf{R}(\mathbf{x_{j-1}})\|_{\mathbf{X}}\,, \quad (22)$$

where $\mathbf{R(z)} = \mathcal{L}\mathbf{z} + \mathbf{b} - \mathbf{z}$ is the *residual* of the functional equation considered.

By applying the general theory, we also obtain a lower bound of the error

$$\|\mathbf{x_j} - \mathbf{x_\odot}\|_{\mathbf{X}} \quad \geq \quad \frac{\mathbf{1}}{\mathbf{1+q}} \|\mathbf{x_{j+1}} - \mathbf{x_j}\|_{\mathbf{X}} \quad = \quad \frac{\mathbf{1}}{\mathbf{1+q}} \|\mathbf{R(x_j)}\|_{\mathbf{X}} \,. \quad (23)$$

Hence, we arrive at the following estimates for the error in the linear operator equation:

$$\frac{\mathbf{1-q}}{\mathbf{q}} \|\mathbf{x_j} - \mathbf{x_\odot}\|_{\mathbf{X}} \leq \|\mathbf{R(x_{j-1})}\|_{\mathbf{X}} \leq (\mathbf{1+q}) \|\mathbf{x_{j-1}} - \mathbf{x_\odot}\|_{\mathbf{X}} \,.$$

Iteration methods in linear algebra

Important applications of the above results are associated with systems of linear simultaneous equations and other algebraic problems. Set $\mathbf{X} = \mathbb{R}^{\mathbf{n}}$ and assume that $\mathcal{L}$ is defined by a nondegenerate matrix $\mathbf{A} \in \mathbb{M}^{\mathbf{n}\times\mathbf{n}}$ decomposed into three matrixes

$$\mathbf{A} = \mathbf{A}_{\ell} + \mathbf{A}_{\mathbf{d}} + \mathbf{A}_{\mathbf{r}},$$

where $\mathbf{A}_{\ell}$, $\mathbf{A}_{\mathbf{r}}$, and $\mathbf{A}_{\mathbf{d}}$ are certain lower, upper, and diagonal matrices, respectively.

Iteration methods for systems of linear simultaneous equations associated with $\mathbf{A}$ are often represented in the form

$$\mathbf{B}\frac{\mathbf{x_i} - \mathbf{x_{i-1}}}{\tau} + \mathbf{A}\,\mathbf{x_{i-1}} = \mathbf{f}\,. \tag{24}$$

In (24), the matrix $\mathbf{B}$ and the parameter τ may be taken in various ways (depending on the properties of $\mathbf{A}$). We consider three frequently encountered cases:

(a) $\mathbf{B} = \mathbf{A_d}$,

(b) $\mathbf{B} = \mathbf{A_d} + \mathbf{A}_\ell$,

(c) $\mathbf{B} = \mathbf{A_d} + \omega\mathbf{A}_\ell$, $\tau = \omega$.

For $\boldsymbol{\tau} = 1$, (a) and (b) lead to the methods of Jacobi and Zeidel, respectively. In (c), the parameter ω must be in the interval $(0, 2)$. If $\omega > 1$, we have the so-called "upper relaxation method", and $\omega < 1$ corresponds to the "lower relaxation method".

The method (24) is reduced to (19) if we set

$$\mathcal{L} = \mathbb{I} - \boldsymbol{\tau}\mathbf{B}^{-1}\mathbf{A} \qquad \text{and} \qquad \mathbf{b} = \boldsymbol{\tau}\mathbf{B}^{-1}\mathbf{f}\,, \tag{25}$$

where $\mathbb{I}$ is the unit matrix. It is known that $\mathbf{x_i}$ converges to $\mathbf{x}_\odot$ that is a solution of the system

$$\mathbf{A}\,\mathbf{x}_\odot = \mathbf{f} \tag{26}$$

if an only if all the eigenvalues of $\mathcal{L}$ are less than one.
Obviously, $\mathbf{B}$ and $\boldsymbol{\tau}$ should be taken in such a way that they guarantee the fulfillment of this condition.

Assume that $\|\mathcal{L}\| \leq \mathbf{q} < \mathbf{1}$. In view of (21)-(23), the quantities

$$\mathbf{M}^{\mathbf{i}}_{\oplus} = \mathbf{q}(\mathbf{1}-\mathbf{q})^{-1} \|\mathbf{R}(\mathbf{x_{i-1}})\|, \tag{27}$$

$$\mathbf{M}^{\mathbf{0i}}_{\oplus} = \mathbf{q}^{\mathbf{i}}(\mathbf{1}-\mathbf{q})^{-1} \|\mathbf{R}(\mathbf{x_0})\|, \tag{28}$$

$$\mathbf{M}^{\mathbf{i}}_{\ominus} = (\mathbf{1}+\mathbf{q})^{-1} \|\mathbf{R}(\mathbf{x_i})\| \tag{29}$$

furnish upper and lower bounds of the error for the vector $\mathbf{x_i}$.

Remark. It is worth noting that from the practical viewpoint finding an upper bound for $\|\mathcal{L}\|$ and proving that it is less than 1 presents a special and often not easy task.
If **q** is very close to 1, then the convergence of an iteration process may be very slow. As we have seen, in this case, the quality of error estimates is also degraded. A well–accepted way for accelerating the convergence consists of using a modified system obtained from the original one by means of a suitable *preconditioner*

Task 2

Consider the problem

$$\mathbf{Ax} = \mathbf{f}$$

for a symmetric matrix $\mathbf{A}$ with coefficients

$$\mathbf{a_{ij}} = \kappa/\mathbf{ij} \quad \text{if } \mathbf{i} \neq \mathbf{j}, \quad \kappa = \mathbf{0.1}$$
$$\mathbf{a_{ii}} = \mathbf{i}.$$

Solved the system by the iteration method

$$\mathbf{x_{i+1}} = (\mathbb{I} - \tau \mathbf{B}^{-1}\mathbf{A})\,\mathbf{x_i} + \tau \mathbf{B}^{-1}\mathbf{F}$$

with $\mathbf{B} = \mathbf{A_D}$ and $\mathbf{x_0} = \{\mathbf{0}, \mathbf{0}, ...\mathbf{0}\}$, determine q and define two–sided error bounds.

Applications to integral equations

Many problems in science and engineering can be stated in terms of integral equations. One of the most typical cases is to find a function $\mathbf{x_\odot(t)} \in \mathbf{C[a,b]}$ such that

$$\mathbf{x_\odot(t)} = \lambda \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K(t,s)}\, \mathbf{x_\odot(s)}\, \mathbf{ds} + \mathbf{f(t)}, \tag{30}$$

where $\lambda \geq 0$, $\mathbf{K}$ (the kernel) is a continuous function for

$$\mathbf{(x,t)} \in \mathbf{Q} := \{\mathbf{a} \leq \mathbf{s} \leq \mathbf{b},\ \mathbf{a} \leq \mathbf{t} \leq \mathbf{b}\}$$

and

$$|\mathbf{K(t,s)}| \leq \mathbf{M}, \qquad \forall \mathbf{(t,s)} \in \mathbf{Q}.$$

Also, we assume that $\mathbf{f} \in \mathbf{C[a,b]}$.

Let us define the operator $\mathfrak{T}$ as follows:

$$\mathbf{y(t)} := \mathfrak{T}\mathbf{x(t)} := \lambda \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K(t,x)x(s)\,ds} + \mathbf{f(t)} \tag{31}$$

and show that $\mathfrak{T}$ maps continuous functions to continuous ones. Let $\mathbf{t_0}$ and $\mathbf{t_0 + \Delta t}$ belong to $\mathbf{[a,b]}$. Then,

$$\begin{aligned} |\mathbf{y(t_0 + \Delta t)} - \mathbf{y(t_0)}| \leq \\ \leq |\lambda| \int_{\mathbf{a}}^{\mathbf{b}} |\mathbf{K(t_0 + \Delta t, s)} - \mathbf{K(t_0, s)}||\mathbf{x(s)}|\,\mathbf{ds} + \\ + |\mathbf{f(t_0 + \Delta t)} - \mathbf{f(t_0)}|. \end{aligned}$$

Since $\mathbf{K}$ and $\mathbf{f}$ are continuous on the compact sets $\mathbf{Q}$ and $\mathbf{[a,b]}$, respectively, they are uniformly continuous on these sets.

Therefore, for any given ε one can find a small number δ such that

$$|\mathbf{f}(\mathbf{t_0} + \mathbf{\Delta t}) - \mathbf{f}(\mathbf{t_0})| < \varepsilon$$

and

$$|\mathbf{K}(\mathbf{t_0} + \mathbf{\Delta t}, \mathbf{s}) - \mathbf{K}(\mathbf{t_0}, \mathbf{s})| < \varepsilon,$$

provided that $|\mathbf{\Delta t}| < \delta$.
Thus, we have

$$|\mathbf{y}(\mathbf{t_0} + \mathbf{\Delta t}) - \mathbf{y}(\mathbf{t_0})| \leq \varepsilon(|\lambda||\mathbf{b} - \mathbf{a}| \max_{\mathbf{s}\in[\mathbf{a},\mathbf{b}]} |\mathbf{x}(\mathbf{s})| + \mathbf{1}) = \mathbf{C}\varepsilon,$$

and, consequently, $\mathbf{y}(\mathbf{t_0} + \mathbf{\Delta t})$ tends to $\mathbf{y}(\mathbf{t_0})$ as $|\mathbf{\Delta t}| \to \mathbf{0}$.

$\mathfrak{T} : \mathbf{C[a,b]} \to \mathbf{C[a,b]}$ is a **contractive mapping**. Indeed,

$$\mathbf{d}(\mathfrak{T}\mathbf{x}, \mathfrak{T}\mathbf{y}) = \max_{\mathbf{a}\le\mathbf{t}\le\mathbf{b}} |\mathfrak{T}\mathbf{x}(\mathbf{t}) - \mathfrak{T}\mathbf{y}(\mathbf{t})| =$$

$$= \max_{\mathbf{a}\le\mathbf{t}\le\mathbf{b}} \left|\lambda \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K}(\mathbf{t},\mathbf{s})(\mathbf{x}(\mathbf{s}) - \mathbf{y}(\mathbf{s}))\, \mathbf{ds}\right| \le$$

$$\le |\lambda| \mathbf{M}(\mathbf{b}-\mathbf{a}) \max_{\mathbf{a}\le\mathbf{s}\le\mathbf{b}} |\mathbf{x}(\mathbf{s}) - \mathbf{y}(\mathbf{s})| = |\lambda| \mathbf{M}(\mathbf{b}-\mathbf{a}) \mathbf{d}(\mathbf{x},\mathbf{y}),$$

so that $\mathfrak{T}$ is a $\mathbf{q}$-contractive operator with

$$\mathbf{q} = |\lambda| \mathbf{M}(\mathbf{b}-\mathbf{a}), \tag{32}$$

provided that

$$|\lambda| < \frac{\mathbf{1}}{\mathbf{M}(\mathbf{b}-\mathbf{a})}. \tag{33}$$

Numerical procedure

An approximate solution of (30) can be found by the iteration method

$$\mathbf{x_{i+1}(t)} = \lambda \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K(t,s)x_i(s)\,ds} + \mathbf{f(t)}. \tag{34}$$

If (33) holds, then from the Banach theorem it follows that the sequence $\{\mathbf{x_i}\}$ converges to the exact solution.
We apply the theory exposed above and find that the accuracy of $\mathbf{x_i}$ is subject to the estimate

$$\frac{\mathbf{1}}{\mathbf{1+q}} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K(t,s)(x_{i+1}(s) - x_i(s))\,ds} \leq$$

$$\leq \max_{\mathbf{a \leq t \leq b}} \mathbf{|x_i(t) - x_\odot(t)|} \leq \frac{\mathbf{q}}{\mathbf{1-q}} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K(t,s)(x_i(s) - x_{i-1}(s))\,ds}. \tag{35}$$

Applications to Volterra type equations

Consider the fixed point problem

$$\mathbf{x}_{\odot}(\mathbf{t}) = \lambda \int_{\mathbf{a}}^{\mathbf{t}} \mathbf{K}(\mathbf{t},\mathbf{s})\,\mathbf{x}_{\odot}(\mathbf{s})\,\mathbf{ds} + \mathbf{f}(\mathbf{t}), \tag{36}$$

where

$$|\mathbf{K}(\mathbf{t},\mathbf{s})| \leq \mathbf{M}, \qquad \forall(\mathbf{t},\mathbf{s}) \in \mathbf{Q}$$

and $\mathbf{f} \in \mathbf{C}[\mathbf{a},\mathbf{b}]$.
Define the operator $\mathbf{T}$ as follows:

$$\mathbf{Tx}(\mathbf{t}) = \lambda \int_{\mathbf{a}}^{\mathbf{t}} \mathbf{K}(\mathbf{t},\mathbf{s})\,\mathbf{x}(\mathbf{s})\,\mathbf{ds} + \mathbf{f}(\mathbf{t}).$$

Similarly, to the previous case we establish that

$$\mathbf{d}(\mathbf{Tx},\mathbf{Ty}) \leq |\lambda|\mathbf{M}(\mathbf{t}-\mathbf{a})\mathbf{d}(\mathbf{x},\mathbf{y}).$$

By the same arguments we find that

$$\mathbf{d}(\mathbf{T}^{\mathbf{n}}\mathbf{x}, \mathbf{T}^{\mathbf{n}}\mathbf{y}) \leq |\lambda|^{\mathbf{n}}\mathbf{M}^{\mathbf{n}} \frac{(\mathbf{t}-\mathbf{a})^{\mathbf{n}}}{\mathbf{n}!}\mathbf{d}(\mathbf{x}, \mathbf{y}),$$

Thus, the operator $\mathfrak{T} := \mathbf{T}^{\mathbf{n}}$ is $\mathbf{q}$-contractive with a certain $\mathbf{q} < \mathbf{1}$, provided that $\mathbf{n}$ is large enough.
In view of Proposition 1, we conclude that the iteration method converges to $\mathbf{x}_{\odot}$ and the errors are controlled by the two–sided error estimates.

Applications to ordinary differential equations

Let $\mathbf{u}$ be a solution of the simplest initial boundary-value problem

$$\frac{\mathbf{du}}{\mathbf{dt}} = \varphi(\mathbf{t}, \mathbf{u}(\mathbf{t})), \quad \mathbf{u}(\mathbf{t_0}) = \mathbf{a}, \tag{37}$$

where the solution $\mathbf{u}(\mathbf{t})$ is to be found on the interval $[\mathbf{t_0}, \mathbf{t_1}]$. Assume that the function $\varphi(\mathbf{t}, \mathbf{p})$ is continuous on the set

$$\mathbf{Q} = \{\mathbf{t_0} \le \mathbf{t} \le \mathbf{t_1},\ \mathbf{a} - \mathbf{\Delta} \le \mathbf{p} \le \mathbf{a} + \mathbf{\Delta}\}$$

and

$$|\varphi(\mathbf{t}, \mathbf{p_1}) - \varphi(\mathbf{t}, \mathbf{p_2})| \le \mathbf{L}|\mathbf{p_1} - \mathbf{p_2}|, \quad \forall (\mathbf{t}, \mathbf{p}) \in \mathbf{Q}. \tag{38}$$

Problem (37) can be reduced to the integral equation

$$\mathbf{u(t)} = \int_{\mathbf{t_0}}^{\mathbf{t}} \varphi(\mathbf{s}, \mathbf{u(s)})\, \mathbf{ds} + \mathbf{a} \tag{39}$$

and it is natural to solve the latter problem by the iteration method

$$\mathbf{u_j(t)} = \int_{\mathbf{t_0}}^{\mathbf{t}} \varphi(\mathbf{s}, \mathbf{u_{j-1}(s)})\, \mathbf{ds} + \mathbf{a}. \tag{40}$$

To justify this procedure, we must verify that the operator

$$\mathfrak{T}\mathbf{u} := \int_{\mathbf{t_0}}^{\mathbf{t}} \varphi(\mathbf{s}, \mathbf{u(s)})\, \mathbf{ds} + \mathbf{a}$$

is **q**-contractive with respect to the norm

$$\|\mathbf{u}\| := \max_{\mathbf{t}\in[\mathbf{t_0},\mathbf{t_1}]} |\mathbf{u(t)}|. \tag{41}$$

We have

$$\|\mathfrak{T}\mathbf{z} - \mathfrak{T}\mathbf{y}\| = \max_{\mathbf{t}\in[\mathbf{t_0},\mathbf{t_1}]} \left| \int_{\mathbf{t_0}}^{\mathbf{t}} (\varphi(\mathbf{s},\mathbf{z}(\mathbf{s})) - \varphi(\mathbf{s},\mathbf{y}(\mathbf{s}))\,\mathbf{ds} \right| \leq$$

$$\leq \max_{\mathbf{t}\in[\mathbf{t_0},\mathbf{t_1}]} \mathbf{L} \int_{\mathbf{t_0}}^{\mathbf{t}} |\mathbf{z}(\mathbf{s}) - \mathbf{y}(\mathbf{s})|\,\mathbf{ds} \leq \mathbf{L} \int_{\mathbf{t_0}}^{\mathbf{t_1}} |\mathbf{z}(\mathbf{s}) - \mathbf{y}(\mathbf{s})|\,\mathbf{ds} \leq$$

$$\leq \mathbf{L}(\mathbf{t_1} - \mathbf{t_0}) \max_{\mathbf{s}\in[\mathbf{t_0},\mathbf{t_1}]} |\mathbf{z}(\mathbf{s}) - \mathbf{y}(\mathbf{s})| = \mathbf{L}(\mathbf{t_1} - \mathbf{t_0})\|\mathbf{z} - \mathbf{y}\|.$$

We see that if

$$\mathbf{t_1} < \mathbf{t_0} + \mathbf{L}^{-1}, \tag{42}$$

then the operator $\mathfrak{T}$ is $\mathbf{q}$-contractive with

$$\mathbf{q} := \mathbf{L}(\mathbf{t_1} - \mathbf{t_0}) < \mathbf{1}.$$

Therefore, if the interval $[\mathbf{t_0}, \mathbf{t_1}]$ is small enough (i.e., it satisfies the condition 42), then the existence and uniqueness of a continuous solution $\mathbf{u(t)}$ follows from the Banach theorem. In this case, the solution can be found by the iteration procedure whose accuracy is explicitly controlled by the two–sided error estimates.

For a more detailed investigation of the fixed point methods for integral and differential equations see

A. N. Kolmogorov and S. V. Fomin. *Introductory real analysis*. Dover Publications, Inc., New York, 1975.

E. Zeidler. *Nonlinear functional analysis and its applications. I. Fixed-point theorems*. Springer-Verlag, New York, 1986.

A posteriori estimates based on monotonicity.

The theory of monotone operators gives another way of constructing a posteriori estimates.

Monotone operators are defined on the so–called ordered (or partially ordered) spaces that introduce the relation $\mathbf{x} \leq \mathbf{y}$ for all (or almost all) elements $\mathbf{x}, \mathbf{y}$ of the space.

Definition

An operator $\mathfrak{T}$ is called monotone if $\mathbf{x} \leq \mathbf{y}$ implies $\mathfrak{T}\mathbf{x} \leq \mathfrak{T}\mathbf{y}$.

Consider the fixed point problem

$$\mathbf{x}_\odot = \mathfrak{T}\mathbf{x}_\odot + \mathbf{f}$$

on an ordered (partially ordered) space X. Assume that

$$\mathfrak{T} = \mathfrak{T}_\oplus + \mathfrak{T}_\ominus,$$

$\mathfrak{T}_\oplus$ is monotone,
$\mathfrak{T}_\ominus$ is antitone: $\mathbf{x} \le \mathbf{y}$ implies $\mathfrak{T}\mathbf{x} \ge \mathfrak{T}\mathbf{y}$,
$\mathfrak{T}_\oplus$ and $\mathfrak{T}_\ominus$ have a common set of images $\mathbf{D}$ which is a convex subset of $\mathbf{X}$.

Next, let $\mathbf{x}_{\ominus 0}$, $\mathbf{x}_{\ominus 1}$, $\mathbf{x}_{\oplus 0}, \mathbf{x}_{\oplus 1} \in \mathbf{D}$ be such elements that

$$\begin{aligned}
&\mathbf{x}_{\ominus 0} \leq \mathbf{x}_{\ominus 1} \leq \mathbf{x}_{\oplus 1} \leq \mathbf{x}_{\oplus 0}, \\
&\mathbf{x}_{\ominus 1} = \mathfrak{T}_{\oplus} \mathbf{x}_{\ominus 0} + \mathfrak{T}_{\ominus} \mathbf{x}_{\oplus 0} + \mathbf{f}, \\
&\mathbf{x}_{\oplus 1} = \mathfrak{T}_{\oplus} \mathbf{x}_{\oplus 0} + \mathfrak{T}_{\ominus} \mathbf{x}_{\ominus 0} + \mathbf{f},
\end{aligned}$$

Then, we observe that

$$\begin{aligned}
\mathbf{x}_{\ominus 2} &= \mathfrak{T}_{\oplus} \mathbf{x}_{\ominus 1} + \mathfrak{T}_{\ominus} \mathbf{x}_{\oplus 1} + \mathbf{f} \geq \mathfrak{T}_{\oplus} \mathbf{x}_{\ominus 0} + \mathfrak{T}_{\ominus} \mathbf{x}_{\oplus 0} + \mathbf{f} = \mathbf{x}_{\ominus 1} \\
\mathbf{x}_{\oplus 2} &= \mathfrak{T}_{\oplus} \mathbf{x}_{\oplus 1} + \mathfrak{T}_{\ominus} \mathbf{x}_{\ominus 1} + \mathbf{f} \leq \mathfrak{T}_{\oplus} \mathbf{x}_{\oplus 0} + \mathfrak{T}_{\ominus} \mathbf{x}_{\ominus 0} + \mathbf{f} = \mathbf{x}_{\oplus 1}.
\end{aligned}$$

By continuing the iterations we obtain elements such that

$$\mathbf{x}_{\ominus\,\mathbf{k}} \leq \mathbf{x}_{\ominus\,(\mathbf{k}+\mathbf{1})} \leq \mathbf{x}_{\oplus\,(\mathbf{k}+\mathbf{1})} \leq \mathbf{x}_{\oplus\,\mathbf{k}}.$$

Then $\mathbf{x} \to \mathfrak{T}\mathbf{x} + \mathbf{f}$ maps $\mathbf{D}$ to itself. If $\mathbf{D}$ is compact, then by the Schauder fixed point theorem $\mathbf{x}_{\odot} \in \mathbf{D}$ exists. Moreover, it is bounded from below and above by the sequences $\{\mathbf{x}_{\ominus\,\mathbf{k}}\}$ and $\{\mathbf{x}_{\oplus\,\mathbf{k}}\}$.

Applications of this method are mainly oriented towards systems of linear simultaneous equations and integral equations
(see L. Collatz. Funktionanalysis und numerische mathematik, Springer-Verlag, Berlin, 1964). For example, consider a system of linear simultaneous equations

$$\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{f}$$

that is supposed to have a unique solution $\mathbf{x}_\odot$. Assume that

$$\mathbf{A} = \mathbf{A}_\oplus - \mathbf{A}_\ominus, \qquad \mathbf{A}_\ominus = \{\mathbf{a}_{\mathbf{ij}}^\ominus\} \in \mathbb{M}^{\mathbf{n}\times\mathbf{n}},$$
$$\mathbf{A}_\oplus = \{\mathbf{a}_{\mathbf{ij}}^\oplus\} \in \mathbb{M}^{\mathbf{n}\times\mathbf{n}}, \quad \mathbf{a}_{\mathbf{ij}}^\ominus \geq \mathbf{0}, \quad \mathbf{a}_{\mathbf{ij}}^\oplus \geq \mathbf{0}.$$

We may partially order the space $\mathbb{R}^n$ by saying that $\mathbf{x} \leq \mathbf{y}$ if and only if $\mathbf{x_i} \leq \mathbf{y_i}$ for $i = 1, 2, \ldots n$. Compute the vectors

$$\mathbf{x}_{\ominus(\mathbf{k}+1)} = \mathbf{A}_\oplus \mathbf{x}_{\ominus\,\mathbf{k}} + \mathbf{A}_\ominus \mathbf{x}_{\oplus\,\mathbf{k}} + \mathbf{f}, \quad \mathbf{x}_{\oplus(\mathbf{k}+1)} = \mathbf{A}_\oplus \mathbf{x}_{\oplus\,\mathbf{k}} + \mathbf{A}_\ominus \mathbf{x}_{\ominus\,\mathbf{k}} + \mathbf{f}.$$

If $\mathbf{x}_{\ominus\,\mathbf{0}} \leq \mathbf{x}_{\ominus\,\mathbf{1}} \leq \mathbf{x}_\odot \leq \mathbf{x}_{\oplus\,\mathbf{1}} \leq \mathbf{x}_{\oplus\mathbf{0}}$, then for all the components of $\mathbf{x}_\odot$ we obtain two–sided estimates

$$\mathbf{x}_{\ominus\,\mathbf{k}}^{(\mathbf{i})} \leq \mathbf{x}_{\ominus\,(\mathbf{k}+1)}^{(\mathbf{i})} \leq \mathbf{x}_\odot^{(\mathbf{i})} \leq \mathbf{x}_{\oplus\,(\mathbf{k}+1)}^{(\mathbf{i})} \leq \mathbf{x}_{\oplus\mathbf{k}}^{(\mathbf{i})}, \quad \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots \mathbf{n}.$$

Task 3.

Apply the above method for finding two–sided bounds of the Euclid error norm and componentwise errors for a system of linear simultaneous equations

$$\mathbf{Ax} = \mathbf{f}$$

where

$$\mathbf{a_{ij}} = (-\mathbf{1})^{\mathbf{i+j}}\kappa/\mathbf{ij} \quad \text{if } \mathbf{i} \neq \mathbf{j}, \quad \kappa = \mathbf{0.1}$$
$$\mathbf{a_{ii}} = \mathbf{i}.$$

For the ith component of the solution determine the lower and upper bounds as follows:

$$\max_{j=0,1,\ldots k+1} \left(\mathbf{x_j^{\ominus}}\right)_i \leq (\mathbf{x}_{\odot})_i \leq \min_{j=0,1,\ldots k+1} \left(\mathbf{x_j^{\oplus}}\right)_i .$$

It should be remarked that convergence of $\mathbf{x}^{(\mathbf{i})}_{\ominus\,\mathbf{k}}$ and $\mathbf{x}^{(\mathbf{i})}_{\oplus\,\mathbf{k}}$ to $\mathbf{x}_{\odot}$ (and the convergence rate) requires a special investigation, which must use specific features of a particular problem.

In principle, a posteriori error estimates based on monotonicity can provide the most informative POINTWISE a posteriori error estimates. Regrettably, the respective theory has not been yet properly investigated.

Lecture 2

The goal of Lecture 2 to give an overview of a posteriori error estimation methods developed for Finite Element approximations in 70th–80th.

Lecture plan

- Mathematical background;
- Residual type error estimates;
  - Basic idea;
  - Estimates in 1D case;
  - Estimates in 2D case;
  - Comments;
- Methods based on post–processing;
- Methods using adjoint problems;

Sobolev spaces with negative indices

Definition

Linear functionals defined on the functions of the space $\overset{\circ}{C}{}^{\infty}(\Omega)$ are called **distributions**. They form the space $\mathcal{D}'(\Omega)$

Value of a **distribution g** on a function φ is $\langle \mathbf{g}, \varphi \rangle$.
Distributions possess an important property:

they have derivatives of any order .

Let $\mathbf{g} \in \mathcal{D}'(\mathbf{\Omega})$, then the quantity $-\langle \mathbf{g}, \frac{\partial \varphi}{\partial x_i} \rangle$ is another linear functional on $\mathcal{D}(\Omega)$. It is viewed as a generalized partial derivative of **g** taken over the i-th variable.

Derivatives of L^q–functions

Any function g from the space $L^q(\Omega)$ $(q \geq 1)$ defines a certain distribution as

$$\langle \mathbf{g}, \varphi \rangle = \int_{\Omega} \mathbf{g}\varphi \, \mathbf{dx}$$

and, therefore, has generalized derivatives of any order. The sets of distributions, which are derivatives of q-integrable functions, are called Sobolev spaces with negative indices.

Definition

The space $W^{-\ell,q}(\Omega)$ is the space of distributions $\mathbf{g} \in \mathcal{D}'(\Omega)$ such that

$$\mathbf{g} = \sum_{|\boldsymbol{\alpha}| \leq \ell} \mathbf{D}^{\boldsymbol{\alpha}} \mathbf{g}_{\boldsymbol{\alpha}} \,,$$

where $\mathbf{g}_{\boldsymbol{\alpha}} \in L^q(\Omega)$.

Spaces $W^{-1,p}(\Omega)$

$W^{-1,p}(\Omega)$ contains distributions that can be viewed as generalized derivatives of L^q-functions. The functional

$$\left\langle \frac{\partial \mathbf{f}}{\partial \mathbf{x_i}}, \varphi \right\rangle := -\int_{\Omega} \mathbf{f} \frac{\partial \varphi}{\partial \mathbf{x_i}} \, \mathbf{dx} \qquad \mathbf{f} \in \mathbf{L^q(\Omega)}$$

is linear and continuous not only for $\varphi \in \overset{\circ}{C}{}^{\infty}(\Omega)$ but, also, for $\varphi \in \overset{\circ}{W}{}^{1,p}(\Omega)$, where $1/p + 1/q = 1$ (density property). Hence, first generalized derivatives of $\mathbf{f}$ lie in the space dual to $\overset{\circ}{\mathbf{W}}{}^{\mathbf{1,p}}(\mathbf{\Omega})$ denoted by $\mathbf{W^{-1,p}(\Omega)}$.

For $\overset{\circ}{\mathbf{W}}{}^{\mathbf{1,2}}(\mathbf{\Omega}) = \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\mathbf{\Omega})$, the respective dual space is denoted by $\mathbf{H^{-1}(\Omega)}$.

Norms in "negative spaces"

For $\mathbf{g} \in H^{-1}(\Omega)$ we may introduce two equivalent "negative norms".

$$\|\mathbf{g}\|_{(-1),\Omega} := \sup_{\varphi \in \overset{\circ}{\mathbf{H}}{}^{1}(\Omega)} \frac{|\langle \mathbf{g}, \varphi \rangle|}{\|\varphi\|_{1,2,\Omega}} < +\infty$$

$$\mathbf{|g|} := \sup_{\varphi \in \overset{\circ}{\mathbf{H}}{}^{1}(\Omega)} \frac{|\langle \mathbf{g}, \varphi \rangle|}{\|\nabla \varphi\|_{\Omega}} < +\infty$$

From the definitions, it follows that

$$\langle \mathbf{g}, \varphi \rangle \leq \|\mathbf{g}\|_{(-1),\Omega} \|\varphi\|_{1,2,\Omega}$$

$$\boxed{\langle \mathbf{g}, \varphi \rangle \leq \mathbf{|g|} \quad \|\nabla \varphi\|_{\Omega}}$$

Errors and Residuals. First glance

If an analyst is not sure in the quality of an approximate solution computed, then the very first idea that comes to his mind is to substitute the approximate solution into the equation and look at the **equation residual**.

We begin by recalling basic relations between residuals and errors that hold for systems of linear simultaneous equations. Let $\mathcal{A} \in \mathbb{M}^{n\times n}$, $\det \mathcal{A} \neq 0$, consider the system

$$\mathcal{A}\mathbf{u} + \mathbf{f} = \mathbf{0}.$$

For any v we have the simplest **residual** type estimate

$$\mathcal{A}(\mathbf{v} - \mathbf{u}) = \mathcal{A}\mathbf{v} + \mathbf{f}; \quad \Rightarrow \quad \|\mathbf{e}\| \leq \|\mathcal{A}^{-1}\|\|\mathbf{r}\|.$$

where $\mathbf{e} = \mathbf{v} - \mathbf{u}$ and $\mathbf{r} = \mathcal{A}\mathbf{v} + \mathbf{f}$.

Two–sided estimates

Define the quantities

$$\lambda_{\min} = \min_{\substack{y \in \mathbb{R}^n \\ y \neq 0}} \frac{\|\mathcal{A}y\|}{\|y\|} \quad \text{and} \quad \lambda_{\max} = \max_{\substack{y \in \mathbb{R}^n \\ y \neq 0}} \frac{\|\mathcal{A}y\|}{\|y\|}$$

Since $\mathcal{A}\mathbf{e} = \mathbf{r}$, we see that

$$\lambda_{\min} \leq \frac{\|\mathcal{A}\mathbf{e}\|}{\|\mathbf{e}\|} = \frac{\|\mathbf{r}\|}{\|\mathbf{e}\|} \leq \lambda_{\max} \;\Rightarrow\; \lambda_{\max}^{-1}\|\mathbf{r}\| \leq \|\mathbf{e}\| \leq \lambda_{\min}^{-1}\|\mathbf{r}\|.$$

Since $\mathbf{u}$ is a solution, we have

$$\lambda_{\min} \leq \frac{\|\mathcal{A}\mathbf{u}\|}{\|\mathbf{u}\|} = \frac{\|\mathbf{f}\|}{\|\mathbf{u}\|} \leq \lambda_{\max} \;\Rightarrow\; \lambda_{\max}^{-1}\|\mathbf{f}\| \leq \|\mathbf{u}\| \leq \lambda_{\min}^{-1}\|\mathbf{f}\|$$

Thus,

$$\frac{\lambda_{\min}}{\lambda_{\max}} \frac{\|\mathbf{r}\|}{\|\mathbf{f}\|} \leq \frac{\|\mathbf{e}\|}{\|\mathbf{u}\|} \leq \frac{\lambda_{\max}}{\lambda_{\min}} \frac{\|\mathbf{r}\|}{\|\mathbf{f}\|}.$$

Key "residual–error" relation

Since

$$\frac{\lambda_{\max}}{\lambda_{\min}} = \mathbf{Cond}\,\mathcal{A},$$

we arrive at the basic relation where the matrix condition number serves as an important factor

$$\boxed{(\mathbf{Cond}\,\mathcal{A})^{-1}\frac{\|\mathbf{r}\|}{\|\mathbf{f}\|} \le \frac{\|\mathbf{e}\|}{\|\mathbf{u}\|} \le \mathbf{Cond}\,\mathcal{A}\,\frac{\|\mathbf{r}\|}{\|\mathbf{f}\|}.} \qquad (43)$$

Thus, the relative error is controlled by the relative value of the residual. However, the bounds deteriorates when the conditional number is large.

In principle, the above consideration can extended to a wider set of linear problems, where

$$\mathcal{A} \in \mathcal{L}(\mathbf{X}, \mathbf{Y})$$

is a coercive linear operator acting from a Banach space $\mathbf{X}$ to another space $\mathbf{Y}$ and $\mathbf{f}$ is a given element of Y.

However, if $\mathcal{A}$ is related to a boundary-value problem, then one should properly define the spaces $\mathbf{X}$ and $\mathbf{Y}$ and find a practically meaningful analog of the estimate (43).

Elliptic equations

Let $\mathcal{A} : \mathbf{X} \longrightarrow \mathbf{Y}$ be a linear elliptic operator. Consider the boundary-value problem

$$\mathcal{A}\mathbf{u} + \mathbf{f} = \mathbf{0} \quad \text{in } \boldsymbol{\Omega}, \qquad \mathbf{u} = \mathbf{u_0} \quad \text{on } \partial\boldsymbol{\Omega}.$$

Assume that $\mathbf{v} \in \mathbf{X}$ is an approximation of $\mathbf{u}$. Then, we should measure the error in $\mathbf{X}$ and the residual in $\mathbf{Y}$, so that the principal form of the estimate is

$$\|\mathbf{v} - \mathbf{u}\|_{\mathbf{X}} \leq \mathbf{C}\|\mathcal{A}\mathbf{v} + \mathbf{f}\|_{\mathbf{Y}}, \tag{44}$$

where the constant $\mathbf{C}$ is independent of $\mathbf{v}$. The key question is as follows:

Which spaces X and Y should we choose for a particular boundary-value problem ?

Consider the problem

$$\mathbf{\Delta u} + \mathbf{f} = \mathbf{0} \quad \text{in} \mathbf{\Omega}, \qquad \mathbf{u} = \mathbf{0} \quad \text{on} \partial\mathbf{\Omega},$$

with $\mathbf{f} \in \mathbf{L^2(\Omega)}$. The generalized solution satisfies the relation

$$\int_{\mathbf{\Omega}} \nabla \mathbf{u} \cdot \nabla \mathbf{w} \, \mathbf{dx} = \int_{\mathbf{\Omega}} \mathbf{f} \mathbf{w} \, \mathbf{dx} \quad \forall \mathbf{w} \in \mathbf{V_0} := \mathring{\mathbf{H}}^{\mathbf{1}}(\mathbf{\Omega}),$$

which implies the **energy estimate**

$$\|\nabla \mathbf{u}\|_{\mathbf{2},\mathbf{\Omega}} \leq \mathbf{C_\Omega} \|\mathbf{f}\|_{\mathbf{2},\mathbf{\Omega}}.$$

Here $\mathbf{C_\Omega}$ is a constant in the Friederichs-Steklov inequality. Assume that an approximation $\mathbf{v} \in \mathbf{V_0}$ and $\mathbf{\Delta v} \in \mathbf{L^2(\Omega)}$. Then,

$$\int_{\mathbf{\Omega}} \nabla(\mathbf{u} - \mathbf{v}) \cdot \nabla \mathbf{w} \, \mathbf{dx} = \int_{\mathbf{\Omega}} (\mathbf{f} + \mathbf{\Delta v}) \mathbf{w} \, \mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

Setting $\mathbf{w} = \mathbf{u} - \mathbf{v}$, we obtain the estimate

$$\|\nabla(\mathbf{u} - \mathbf{v})\|_{2,\Omega} \leq \mathbf{C}_\Omega \|\mathbf{f} + \boldsymbol{\Delta}\mathbf{v}\|_{2,\Omega}, \tag{45}$$

whose right-hand side of (45) is formed by the $\mathbf{L}^2$-norm of the residual. However, usually a sequence of approximations $\{\mathbf{v_k}\}$ converges to $\mathbf{u}$ only in the energy space, i.e.,

$$\{\mathbf{v_k}\} \rightarrow \mathbf{u} \qquad \text{in } \mathbf{H}^1(\Omega),$$

so that $\|\boldsymbol{\Delta}\mathbf{v_k} + \mathbf{f}\|$ may not converge to zero !

This means that the consistency (the key property of any practically meaningful estimate) is lost.

Which norm of the residual leads to a consistent estimate of the error in the energy norm?

To find it, we should consider Δ not as $\mathbf{H^2} \longrightarrow \mathbf{L^2}$ mapping, but as $\mathbf{H^1} \longrightarrow \mathbf{H^{-1}}$ mapping. For this purpose we use the integral identity

$$\int_\Omega \nabla \mathbf{u} \cdot \nabla \mathbf{w} \, \mathbf{dx} = \langle \mathbf{f}, \mathbf{w} \rangle, \quad \forall \, \mathbf{w} \in \mathbf{V_0} := \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\Omega).$$

Here, $\nabla \mathbf{u} \in \mathbf{L^2}$, so that it has derivatives in $\mathbf{H^{-1}}$ and we consider the above as equivalence of two distributions on all trial functions $\mathbf{w} \in V_0$.
By $\langle \mathbf{f}, \mathbf{w} \rangle \leq \| \mathbf{f} \|_{-1} \|\nabla \mathbf{w}\|_{2,\Omega}$, we obtain another "energy estimate"

$$\|\nabla \mathbf{u}\|_{2,\Omega} \leq \| \mathbf{f} \|_{-1}.$$

Consistent residual estimate

Let $\mathbf{v} \in \mathbf{V_0}$ be an approximation of $\mathbf{u}$. We have

$$\int_\Omega \nabla(\mathbf{u}-\mathbf{v})\cdot\nabla\mathbf{w}\,\mathbf{dx} = \int_\Omega (\mathbf{fw}-\nabla\mathbf{v}\cdot\nabla\mathbf{w})\,\mathbf{dx} =$$

$$= \langle \boldsymbol{\Delta}\mathbf{v}+\mathbf{f}, \mathbf{w}\rangle, \quad \mathbf{f}+\boldsymbol{\Delta}\mathbf{v} \in \mathbf{H}^{-1}(\boldsymbol{\Omega}).$$

By setting $\mathbf{w} = \mathbf{v}-\mathbf{u}$, we obtain

$$\|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega} \le \rceil\!\rceil\, \mathbf{f}+\boldsymbol{\Delta}\mathbf{v} \,\lceil\!\lceil . \tag{46}$$

where

$$\rceil\!\rceil\, \mathbf{f}+\boldsymbol{\Delta}\mathbf{v} \,\lceil\!\lceil = \sup_{\varphi\in\mathring{\mathbf{H}}^{\mathbf{1}}(\boldsymbol{\Omega})} \frac{|\,\langle \mathbf{f}+\boldsymbol{\Delta}\mathbf{v}, \varphi\rangle\,|}{\|\nabla\varphi\|} =$$

$$= \sup_{\varphi\in\mathring{\mathbf{H}}^{\mathbf{1}}(\boldsymbol{\Omega})} \frac{|\int_\Omega \nabla(\mathbf{u}-\mathbf{v})\cdot\nabla\varphi\,|}{\|\nabla\varphi\|} \le \sup_{\varphi\in\mathring{\mathbf{H}}^{\mathbf{1}}(\boldsymbol{\Omega})} \frac{\|\nabla(\mathbf{u}-\mathbf{v})\|\|\nabla\varphi\|}{\|\nabla\varphi\|} \le \|\nabla(\mathbf{u}-\mathbf{v})\|$$

Thus, for the problem considered

$$\|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega} = \mathbf{[}\, \mathbf{f}+\mathbf{\Delta v}\, \mathbf{]} \quad !!! \tag{47}$$

From (47), it readily follows that

$$\mathbf{[}\, \mathbf{f}+\mathbf{\Delta v_k}\, \mathbf{]} \rightarrow \mathbf{0} \quad \text{as} \quad \{\mathbf{v_k}\} \rightarrow \mathbf{u} \text{ in } \mathbf{H^1}.$$

We observe that the estimate (47) is **consistent**.

Diffusion equation

Similar estimates can be derived for

$$\mathcal{A}\mathbf{u} + \mathbf{f} = \mathbf{0}, \quad \text{in } \Omega, \qquad \mathbf{u} = \mathbf{0} \text{ on } \partial\Omega,$$

where

$$\mathcal{A}\mathbf{u} = \mathbf{div}\, \mathbf{A}\nabla\mathbf{u} := \sum_{\mathbf{i},\mathbf{j}=\mathbf{1}}^{\mathbf{d}} \frac{\partial}{\partial \mathbf{x_i}} \left(\mathbf{a_{ij}}(\mathbf{x}) \frac{\partial \mathbf{u}}{\partial \mathbf{x_j}} \right),$$

$$\mathbf{a_{ij}}(\mathbf{x}) = \mathbf{a_{ji}}(\mathbf{x}) \in \mathbf{L}^{\infty}(\Omega),$$

$$\lambda_{\min} |\eta|^2 \leq \mathbf{a_{ij}}(\mathbf{x}) \eta_i \eta_j \leq \lambda_{\max} |\eta|^2, \quad \forall \eta \in \mathbb{R}^{\mathbf{n}}, \ \mathbf{x} \in \Omega,$$

$$\lambda_{\max} \geq \lambda_{\min} \geq 0.$$

Let $\mathbf{v} \in \mathbf{V_0}$ be an approximation of $\mathbf{u}$. Then,

$$\int_\Omega \mathbf{A}\nabla(\mathbf{u}-\mathbf{v})\cdot\nabla\mathbf{w}\,\mathbf{dx} = \int_\Omega (\mathbf{f}\mathbf{w} - \mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{w})\,\mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

Again, the right-hand side of this relation is a bounded linear functional on $\mathbf{V_0}$, i.e.,

$$\mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}) \in \mathbf{H}^{-1}.$$

Hence, we have the relation

$$\int_\Omega \mathbf{A}\nabla(\mathbf{u}-\mathbf{v})\cdot\nabla\mathbf{w}\,\mathbf{dx} = \langle \mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}), \mathbf{w}\rangle, \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

Setting $\mathbf{w} = \mathbf{u} - \mathbf{v}$, we derive the estimate

$$\|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega} \le \lambda_{\min}^{-1} \left[\!\left[\mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}) \right]\!\right]. \tag{48}$$

Next,

$$\| \mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}) \| = \sup_{\varphi \in \mathring{\mathbf{H}}^1(\Omega)} \frac{| \langle \mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}), \varphi \rangle |}{\|\nabla\varphi\|_{2,\Omega}} =$$

$$= \sup_{\varphi \in \mathring{\mathbf{H}}^1(\Omega)} \frac{| \int_\Omega \mathbf{A}\nabla(\mathbf{u}-\mathbf{v}) \cdot \nabla\varphi \, \mathbf{dx} |}{\|\nabla\varphi\|_{2,\Omega}} \le \lambda_{\max} \|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega}. \quad (49)$$

Combining (48) and (49) we obtain

$$\boxed{\lambda_{\max}^{-1} \| \mathbf{R}(\mathbf{v}) \| \le \|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega} \le \lambda_{\min}^{-1} \| \mathbf{R}(\mathbf{v}) \|,} \quad (50)$$

where $\mathbf{R}(\mathbf{v}) = \mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}) \in \mathbf{H}^{-1}(\Omega)$. We see that upper and lower bounds of the error can be evaluated in terms of the negative norm of $\mathbf{R}(\mathbf{v})$.

Main goal

We observe that to find guaranteed bounds of the error reliable estimates of $\|R(v)\|$ are required.

In essence, a posteriori error estimates derived in 70-90' for Finite Element Methods (FEM) offer several approaches to the evaluation of $\|R(v)\|$.
We consider them starting with the so–called **explicit residual method** where such estimates are obtained with help of two key points:

- Galerkin orthogonality property;
- $H^1 \longrightarrow V_h$ interpolation estimates by Clément.

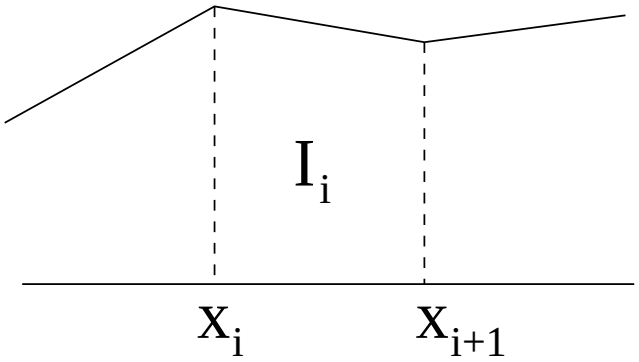
Explicit residual method in 1D case

Take the simplest model

$$(\alpha u')' + f = 0, \qquad u(0) = u(1).$$

Let $I := (0,1)$, $f \in L^2(I)$, $\alpha(x) \in C(\bar{I}) \geq \alpha_0 > 0$. Divide I into a number of subintervals $I_i = (x_i, x_{i+1})$, where $x_0 = 0$, $x_{N+1} = 1$, and $|x_{i+1} - x_i| = h_i$.

Assume that $v \in \overset{\circ}{H}{}^1(I)$ and it is smooth on any interval I_i.



In this case,

$$\| R(v) \| = \sup_{w \in V_0(I),\ w \neq 0} \frac{\int_0^1 (-\alpha v' w' + f w) dx}{\|w'\|_{2,I}} =$$

$$= \sup_{w \in \mathring{H}^1(I);\, w \neq 0} \frac{\sum_{i=0}^N \int_{I_i} (-\alpha v' w' + f w) dx}{\|w'\|_{2,I}} =$$

$$= \sup_{w \in V_0(I),\ w \neq 0} \frac{\sum_{i=0}^N \int_{I_i} r_i(v) w \, dx + \sum_{i=1}^N \alpha(x_i) w(x_i) j(v'(x_i))}{\|w'\|_{2,I}},$$

where $j(\phi(x)) := \phi(x+0) - \phi(x-0)$ is the "jump–function" and $r_i(v) = (\alpha v')' + f$ is the residual on I_i.

For arbitrary v we can hardly get an upper bound for this supremum.

Use Galerkin orthogonality

Assume that $\mathbf{v} = \mathbf{u_h}$, i.e., it is the *Galerkin approximation* obtained on a finite–dimensional subspace $\mathbf{V_{0h}}$ formed by piecewise polynomial continuous functions. Since

$$\int_{\mathbf{I}} \alpha \mathbf{u_h'} \mathbf{w_h'} \, \mathbf{dx} - \int_{\mathbf{I}} \mathbf{f} \mathbf{w_h} \, \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{w_h} \in \mathbf{V_{0h}}.$$

we may add the left–hand side with any w_h to the numerator what gives

$$\| \, \mathbf{R}(\mathbf{u_h}) \, \| = \sup_{\mathbf{w} \in \mathbf{V_0(I)}} \frac{\int_0^1 (-\alpha \mathbf{u_h'} (\mathbf{w} - \pi_\mathbf{h} \mathbf{w})' + \mathbf{f}(\mathbf{w} - \pi_\mathbf{h} \mathbf{w})) \, \mathbf{dx}}{\|\mathbf{w}'\|_{\mathbf{2,I}}},$$

where $\pi_\mathbf{h} : \mathbf{V_0} \to \mathbf{V_{0h}}$ is the interpolation operator defined by the conditions $\pi_\mathbf{h} \mathbf{v} \in \mathbf{V_{0h}}$, $\pi_\mathbf{h} \mathbf{v}(\mathbf{0}) = \pi_\mathbf{h} \mathbf{v}(\mathbf{1}) = \mathbf{0}$ and

$$\pi_\mathbf{h} \mathbf{v}(\mathbf{x_i}) = \mathbf{v}(\mathbf{x_i}), \quad \forall \mathbf{x_i}, \ \mathbf{i} = \mathbf{1}, \mathbf{2}, ..., \mathbf{N}.$$

Integrating by parts

Now, we have

$$\| R(u_h) \| = \sup_{w \in V_0(I)} \left\{ \frac{\sum_{i=0}^{N} \int_{I_i} r_i(u_h)(w - \pi_h w)\, dx}{\|w'\|_{2,I}} + \right.$$
$$\left. + \frac{\sum_{i=1}^{N} \alpha(x_i)(w(x_i) - \pi_h w(x_i)) j(u_h'(x_i))}{\|w'\|_{2,I}} \right\}.$$

Since $w(x_i) - \pi_h w(x_i) = 0$, the second sum vanishes. For first one we have

$$\sum_{i=0}^{N} \int_{I_i} r_i(u_h)(w - \pi_h w)\, dx \le \sum_{i=0}^{N} \|r_i(u_h)\|_{2,I_i} \|w - \pi_h w\|_{2,I_i}.$$

Since for $w \in \overset{\circ}{H}{}^1(I_i)$

$$\|w - \pi_h w\|_{2,I_i} \le c_i \|w'\|_{2,I_i},$$

we obtain for the numerator of the above quotient

$$\sum_{i=0}^{N} \int_{I_i} r_i(u_h)(w - \pi_h w)\, dx \le \sum_{i=0}^{N} c_i \|r_i(u_h)\|_{2,I_i} \|w'\|_{2,I_i} \le$$

$$\le \left(\sum_{i=0}^{N} c_i^2 \|r_i(u_h)\|_{2,I_i}^2 \right)^{1/2} \|w'\|_{2,I},$$

which implies the desired upper bound

$$\boxed{\| R(u_h) \| \le \left(\sum_{i=0}^{N} c_i^2 \|r_i(u_h)\|_{2,I_i}^2 \right)^{1/2}.} \qquad (51)$$

This bound is the sum of local residuals $r_i(u_h)$ with weights given by the **interpolation constants c_i**.

Interpolation constants

For piecewise affine approximations, the interpolation constants c_i are easy to find. Indeed, let γ_i be a constant that satisfies the condition

$$\inf_{w \in \overset{\circ}{H}{}^1(l_i)} \frac{\|w'\|^2_{2,l_i}}{\|w - \pi_h w\|^2_{2,l_i}} \geq \gamma_i .$$

Then, for all $w \in \overset{\circ}{H}{}^1(l_i)$, we have

$$\|w - \pi_h w\|_{2,l_i} \leq \gamma_i^{-1/2} \|w'\|_{2,l_i}$$

and one can set $c_i = \gamma_{l_i}^{-1/2}$.

Let us estimate γ_{l_i}.

Note that

$$\int_{x_i}^{x_{i+1}} |w'|^2 \, dx = \int_{x_i}^{x_{i+1}} |(w - \pi_h w)' + (\pi_h w)'|^2 \, dx,$$

where $(\pi_h w)'$ is constant on (x_i, x_{i+1}). Therefore,

$$\int_{x_i}^{x_{i+1}} (w - \pi_h w)'(\pi_h w)' \, dx = 0$$

and

$$\int_{x_i}^{x_{i+1}} |w'|^2 \, dx = \int_{x_i}^{x_{i+1}} |(w - \pi_h w)'|^2 \, dx + \int_{x_i}^{x_{i+1}} |(\pi_h w)'|^2 \, dx \geq$$
$$\geq \int_{x_i}^{x_{i+1}} |(w - \pi_h w)'|^2 \, dx.$$

Interpolation constants in 1D problem

Thus, we have

$$\inf_{w\in\overset{\circ}{H}{}^1(I_i)} \frac{\int_{x_i}^{x_{i+1}} |w'|^2\,dx}{\int_{x_i}^{x_{i+1}} |w-\pi_h w|^2\,dx} \geq \inf_{w\in\overset{\circ}{H}{}^1(I_i)} \frac{\int_{x_i}^{x_{i+1}} |(w-\pi_h w)'|^2\,dx}{\int_{x_i}^{x_{i+1}} |w-\pi_h w|^2\,dx} \geq$$

$$\geq \inf_{\eta\in\overset{\circ}{H}{}^1(I_i)} \frac{\int_{x_i}^{x_{i+1}} |\eta'|^2\,dx}{\int_{x_i}^{x_{i+1}} |\eta|^2\,dx} = \frac{\pi^2}{h_i^2},$$

so that $\gamma_i = \pi^2/h_i^2$ and $c_i = h_i/\pi$.

Remark. To prove the very last relation we note that

$$\inf_{\eta\in\overset{\circ}{H}{}^1((0,h))} \frac{\int_0^h |\eta'|^2\,dx}{\int_0^h |\eta|^2\,dx} = \frac{\pi^2}{h^2}$$

is attained on the eigenfunction $sin\frac{\pi}{h}x$, of the problem $\phi'' + \lambda\phi = 0$ on $(0, h)$.

Task 4

Solve a boundary–value problem

$$(\alpha \mathbf{v}')' = \mathbf{f},$$
$$\mathbf{v}(\mathbf{0}) = \mathbf{a}, \quad \mathbf{v}(\mathbf{1}) = \mathbf{b}$$

with certain $\alpha(x) > 0$, $\mathbf{f}$, $\mathbf{a}$, and $\mathbf{b}$ by the finite element method with uniform elements (i.e., $\mathbf{h} = \mathbf{1}/\mathbf{N}$). Apply the residual method and compare the errors computed with the true error distribution.

Residual method in 2D case

Let Ω be represented as a union $\mathcal{T}_h$ of simplexes $\mathbf{T_i}$. For the sake of simplicity, assume that $\overline{\Omega} = \cup_{i=1}^{N} \overline{\mathbf{T}}_i$ and $\mathbf{V_{0h}}$ consists of piecewise affine continuous functions. Then the Galerkin approximation $\mathbf{u_h}$ satisfies the relation

$$\int_{\Omega} \mathbf{A}\nabla \mathbf{u_h} \cdot \nabla \mathbf{w_h}\, \mathbf{dx} = \int_{\Omega} \mathbf{f w_h}\, \mathbf{dx}, \quad \forall \mathbf{w_h} \in \mathbf{V_{0h}},$$

where

$$\mathbf{V_{0h}} = \{\mathbf{w_h} \in \mathbf{V_0} \mid \mathbf{w_h} \in \mathbf{P^1}(\mathbf{T_i}),\ \mathbf{T_i} \in \mathcal{F}_\mathbf{h}\}.$$

In this case, negative norm of the residual is

$$\| \mathbf{R}(\mathbf{u_h}) \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{fw} - \mathbf{A}\nabla \mathbf{u_h} \cdot \nabla \mathbf{w})\, \mathbf{dx}}{\|\nabla \mathbf{w}\|_{2,\Omega}}.$$

Let $\pi : \overset{\circ}{H}{}^1 \longrightarrow V_{0h}$ be a continuous interpolation operator. Then, for the **Galerkin approximation**

$$\| \mathbf{R}(\mathbf{u_h}) \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{f}(\mathbf{w} - \pi_\mathbf{h}\mathbf{w}) - \mathbf{A}\nabla \mathbf{u_h} \cdot \nabla (\mathbf{w} - \pi_\mathbf{h}\mathbf{w}))\, \mathbf{dx}}{\|\nabla \mathbf{w}\|_{2,\Omega}}.$$

For finite element approximations such a type projection operators has been constructed. One of the most known was suggested in
Ph. Clément. Approximations by finite element functions using local regularization, *RAIRO Anal. Numér.*, 9(1975).
and is often called the **Clement's interpolation operator**. Its properties play an important role in the a posteriori error estimation method considered.

Clement's Interpolation operator

Let $\mathbf{E_{ij}}$ denote the common edge of the simplexes $\mathbf{T_i}$ and $\mathbf{T_j}$. If $\mathbf{s}$ is an inner node of the triangulation $\mathcal{F}_\mathbf{h}$, then $\omega_\mathbf{s}$ denotes the set of all simplexes having this node.
For any $\mathbf{s}$, we find a polynomial $\mathbf{p_s(x)} \in \mathbf{P^1}(\omega_\mathbf{s})$ such that

$$\int_{\omega_\mathbf{s}} (\mathbf{v} - \mathbf{p_s})\mathbf{q}\,\mathbf{dx} = \mathbf{0} \quad \forall \mathbf{q} \in \mathbf{P^1}(\omega_\mathbf{s}).$$

Now, the interpolation operator $\boldsymbol{\pi_h}$ is defined by setting

$$\pi_\mathbf{h}\mathbf{v}(\mathbf{x_s}) = \mathbf{p}(\mathbf{x_s}), \quad \forall \mathbf{x_s} \in \boldsymbol{\Omega},$$
$$\pi_\mathbf{h}\mathbf{v}(\mathbf{x_s}) = \mathbf{0}, \quad \forall \mathbf{x_s} \in \partial\boldsymbol{\Omega}.$$

It is a linear and continuous mapping of $\overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\boldsymbol{\Omega})$ to the space of piecewise affine continuous functions.

Interpolation estimates in 2D

Moreover, it is subject to the relations

$$\|\mathbf{v} - \pi_h \mathbf{v}\|_{2,\mathbf{T_i}} \leq c_i^T \, \mathbf{diam}\,(\mathbf{T_i}) \|\mathbf{v}\|_{1,2,\omega_N(\mathbf{T_i})}, \tag{52}$$

$$\|\mathbf{v} - \pi_h \mathbf{v}\|_{2,\mathbf{E_{ij}}} \leq c_{ij}^E |\mathbf{E_{ij}}|^{1/2} \|\mathbf{v}\|_{1,2,\omega_E(\mathbf{T_i})}, \tag{53}$$

where $\omega_N(\mathbf{T_i})$ is the union of all simplexes having at least *one common node* with $\mathbf{T_i}$ and $\omega_E(\mathbf{T_i})$ is the union of all simplexes having *a common edge* with $\mathbf{T_i}$.

Interpolation constants c_i^T and c_{ij}^E are LOCAL and depend on the shape of patches $\omega_N(\mathbf{T_i})$ and $\omega_E(\mathbf{T_i})$.

Quotient relations for the constants

Evaluation of c_i^T and c_{ij}^E requires finding *exact lower bounds* of the following variational problems:

$$\gamma_i^T := \inf_{w \in V_0} \frac{\|w\|_{1,2,\omega_N(T_i)}}{\|w - \pi_h w\|_{2,T_i}} \operatorname{diam}(T_i)$$

and

$$\gamma_{ij}^E := \inf_{w \in V_0} \frac{\|w\|_{1,2,\omega_E(T_i)}}{\|w - \pi_h w\|_{2,E_{ij}}} |E_{ij}|^{1/2}.$$

Certainly, we can replace V_0 be $H^1(\omega_N(T_i))$ and $H^1(\omega_E(T_i))$, respectively, but, anyway finding the constants amounts solving functional eigenvalue type problems !

Let $\boldsymbol{\sigma_h} = \mathbf{A}\nabla \mathbf{u_h}$. Then,

$$\| \mathbf{R}(\mathbf{u_h}) \| = \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\int_\Omega (\mathbf{f}(\mathbf{w} - \pi_\mathbf{h}\mathbf{w}) - \boldsymbol{\sigma_h} \cdot \nabla(\mathbf{w} - \pi_\mathbf{h}\mathbf{w}))\, \mathbf{dx}}{\|\nabla \mathbf{w}\|_{2,\Omega}}.$$

If $\boldsymbol{\nu_{ij}}$ is the unit outward normal to $\mathbf{E_{ij}}$, then

$$\int_{\mathbf{T_i}} \boldsymbol{\sigma_h} \cdot \nabla(\mathbf{w} - \pi_\mathbf{h}\mathbf{w})\, \mathbf{dx} =$$

$$= \sum_{\mathbf{E_{ij}} \subset \partial \mathbf{T_i}} \int_{\mathbf{E_{ij}}} (\boldsymbol{\sigma_h} \cdot \boldsymbol{\nu})(\mathbf{w} - \pi_\mathbf{h}\mathbf{w}) \mathbf{ds} - \int_{\mathbf{T_i}} \mathbf{div}\, \boldsymbol{\sigma_h} (\mathbf{w} - \pi_\mathbf{h}\mathbf{w})\, \mathbf{dx},$$

Since on the boundary $\mathbf{w} - \pi_\mathbf{h}\mathbf{w} = \mathbf{0}$, we obtain

$$\| \mathbf{R}(\mathbf{u_h}) \| = \sup_{\mathbf{w}\in\mathbf{V_0}} \left\{ \frac{\sum_{\mathbf{i}=1}^{\mathbf{N}} \int_{\mathbf{T_i}} (\mathbf{div}\, \boldsymbol{\sigma_h} + \mathbf{f})(\mathbf{w} - \pi_\mathbf{h}\mathbf{w})\, \mathbf{dx}}{\|\nabla \mathbf{w}\|_{2,\Omega}} + \right.$$

$$\left. + \frac{\sum_{\mathbf{i}=1}^{\mathbf{N}} \sum_{\mathbf{j}>\mathbf{i}}^{\mathbf{N}} \int_{\mathbf{E_{ij}}} \mathbf{j}(\boldsymbol{\sigma_h} \cdot \boldsymbol{\nu_{ij}})(\mathbf{w} - \pi_\mathbf{h}\mathbf{w})\, \mathbf{ds}}{\|\nabla \mathbf{w}\|_{2,\Omega}} \right\}.$$

First term in sup

$$\int_{T_i} (\mathbf{div}\sigma_h + f)(w - \pi_h w)dx \leq \|\mathbf{div}\sigma_h + f\|_{2,T_i} \|w - \pi_h w\|_{2,T_i}$$
$$\leq c_i^T \|\mathbf{div}\sigma_h + f\|_{2,T_i} \mathbf{diam}\,(T_i) \|w\|_{1,2,\omega_N(T_i)},$$

Then, the first sum is estimated as follows:

$$\sum_{i=1}^{N} \int_{T_i} (\mathbf{div}\,\sigma_h + f)(w - \pi_h w)dx \leq$$
$$\leq d_1 \left(\sum_{i=1}^{N} \left(c_i^T \right)^2 \mathbf{diam}\,(T_i)^2 \|\mathbf{div}\,\sigma_h + f\|_{2,T_i}^2 \right)^{1/2} \|w\|_{1,2,\Omega},$$

where the constant d_1 depends on the maximal number of elements in the set $\omega_N(T_i)$.

Second term in sup

For the second one, we have

$$\sum_{i=1}^{N}\sum_{j>i}^{N}\int_{E_{ij}} j(\sigma_h\cdot\nu_{ij})(w-\pi_h w)\,dx \le$$

$$\le \sum_{i=1}^{N}\sum_{j>i}^{N} \|j(\sigma_h\cdot\nu_{ij})\|_{2,E_{ij}}\; c_{ij}^{E}\; |E_{ij}|^{1/2}\; \|w\|_{1,2,\omega_E(T_i)} \le$$

$$\le d_2\left(\sum_{i=1}^{N}\sum_{j>i}^{N}\left(c_{ij}^{E}\right)^2 |E_{ij}|\|j(\sigma_h\cdot\nu_{ij})\|_{2,E_{ij}}^2\right)^{1/2}\|w\|_{1,2,\Omega},$$

where d_2 depends on the maximal number of elements in the set $\omega_E(T_i)$.

Residual type error estimate

By the above estimates we obtain

$$\| R(u_h) \| \le C_0 \left(\left(\sum_{i=1}^{N} \left(c_i^T \right)^2 \operatorname{diam}(T_i)^2 \| \operatorname{div} \sigma_h + f \|^2_{2,T_i} \right)^{1/2} + \right.$$
$$\left. + \left(\sum_{i=1}^{N} \sum_{j>i}^{N} \left(c_{ij}^E \right)^2 |E_{ij}| \, \| j(\sigma_h \cdot \nu_{ij}) \|^2_{2,E_{ij}} \right)^{1/2} \right). \quad (54)$$

Here $C_0 = C_0(d_1, d_2)$. We observe that the right-hand side is the sum of local quantities (usually denoted by $\eta(T_i)$) multiplied by constants depending on properties of the chosen splitting $\mathcal{F}_h$.

Error indicator for quasi-uniform meshes

For quasi–uniform meshes all generic constants c_i^T have approximately the same value and can be replaced by a single constant c_1. If the constants c_{ij}^E are also estimated by a single constant c_2, then we have

$$\| R(u_h) \| \le C \left(\sum_{i=1}^{N} \eta^2(T_i) \right)^{1/2}, \tag{55}$$

where $C = C(c_1, c_2, C_0)$ and

$$\eta^2(T_i) = c_1^2 \mathrm{diam}\,(T_i)^2 \| \mathrm{div}\,\sigma_h + f \|^2_{2,T_i} + \frac{c_2^2}{2} \sum_{E_{ij} \subset \partial T_i} |E_{ij}| \| j(\sigma_h \cdot \nu_{ij}) \|^2_{2,E_{ij}}.$$

The multiplier $1/2$ arises, because any interior edge is common for two elements.

Comment 1

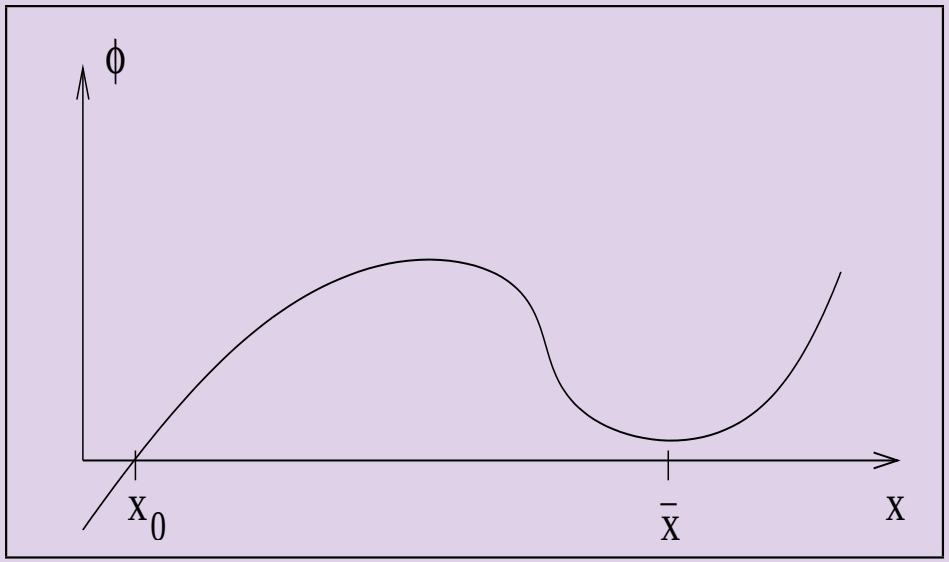
General form of the residual type a posteriori error estimates is as follows:

$$\|\mathbf{u} - \mathbf{u_h}\| \leq \mathbf{M}(\mathbf{u_k}, \mathbf{c_1}, \mathbf{c_2}, \ldots \mathbf{c_N}, \mathcal{D}),$$

where $\mathcal{D}$ is the data set, $\mathbf{u_h}$ is the **Galerkin approximation**, and $\mathbf{c_i}, \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots \mathbf{N}$ are the **interpolation constants**. The constants depend on the **mesh** and properties of the special type interpolation operator. The number $\mathbf{N}$ depends on the dimension of $\mathbf{V_h}$ and may be rather large. If the constants are not sharply defined, then this functional is not more than a certain error indicator. However, in many cases it successfully works and was used in numerous researches.

Comment 2

It is worth noting that for nonlinear problems the dependence between the error and the respective residual is much more complicated. A simple example below shows that the value of the residual may fail to control the distance to the exact solution.



References

It is commonly accepted that this approach brings its origin from the papers
I. Babuska and W. C. Rheinboldt. A-posteriori error estimates for the finite element method. Internat. J. Numer. Meth. Engrg., 12(1978).
I. Babuska and W. C. Rheinboldt. Error estimates for adaptive finite element computations. SIAM J.Numer. Anal., 15(1978). Detailed mathematical analysis of this error estimation method can be found in R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques* Wiley and Sons, Teubner, New-York, 1996.

Also, we recommend the books

M. Ainsworth and T. Oden. *A posteriori error estimation in finite element analysis*, Wiley and Sons, New York, 2000.

K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational differential equations*, Cambridge University Press, Cambridge, 1996

I. Babuska and T. Strouboulis, *The finite element method and its reliability*, Oxford University Press, New York, 2001.

A posteriori methods based on post–processing

Post–processing of approximate solutions is a numerical procedure intended to modify already computed solution in such a way that the post–processed function would fit some a priori known properties much better than the original one.

Preliminaries

Let $\mathbf{e}$ denotes the **error** of an approximate solution $\mathbf{v} \in \mathbf{V}$ and
$\mathcal{E}(\mathbf{v}) : \mathbf{V} \longrightarrow \mathbf{R}_+$ denotes the value of an **error estimator** computed on $\mathbf{v}$.

Definition

The estimator is said to be equivalent to the error for the approximations $\mathbf{v}$ from a certain subset $\widetilde{\mathbf{V}}$ if

$$\mathbf{c_1}\mathcal{E}(\mathbf{v}) \leq \|\mathbf{e}\| \leq \mathbf{c_2}\mathcal{E}(\mathbf{v}) \qquad \forall \mathbf{v} \in \widetilde{\mathbf{V}}$$

Definition

The ratio

$$\mathbf{i_{eff}} := \mathbf{1} + \frac{\mathcal{E}(\mathbf{v}) - \|\mathbf{e}\|}{\|\mathbf{e}\|}$$

is called the **effectivity index** of the estimator $\mathcal{E}$.

Ideal estimator has $\mathbf{i_{eff}} = \mathbf{1}$. However, in real life situations it is hardly possible, so that values $\mathbf{i_{eff}}$ in the diapason from 1 to 2-3 are considered as quite good.

In FEM methods with mesh size **h** one other term is often used:

Definition

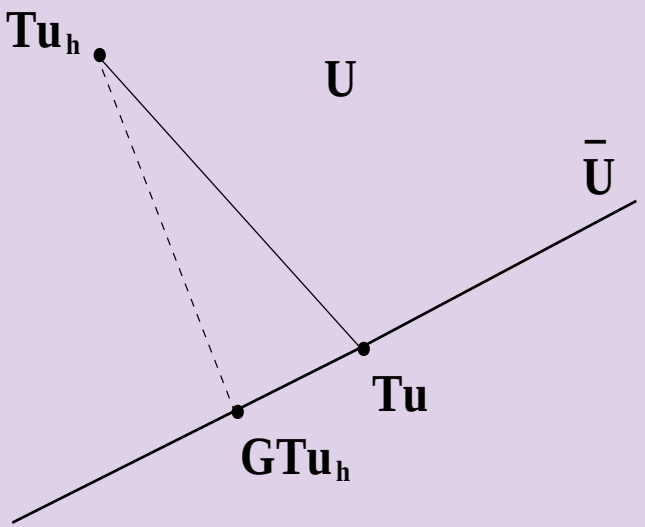
The estimator $\mathcal{E}$ is called **asymptotically equivalent to the error** if for a sequence of approximate solutions $\{u_h\}$ obtained on consequently refined meshes there holds the relation

$$\inf_{\mathbf{h}\to 0} \frac{\mathcal{E}(\mathbf{u_h})}{\|\mathbf{u}-\mathbf{u_h}\|} = \mathbf{1}$$

It is clear that an estimator may be asymptotically exact for one sequence of approximate solutions (e.g. computed on regular meshes) and not exact for another one.

General outlook

Typically, the function $\mathrm{T}\mathbf{u_h}$ (where T is a certain linear operator, e.g., ∇) lies in a space $\mathbf{U}$ that is wider than the space $\bar{\mathbf{U}}$ that contains $\mathrm{T}\mathbf{u}$. If we have a computationally inexpensive continuous mapping $\mathbb{G}$ such that $\mathbb{G}(\mathrm{T}\mathbf{v_h}) \in \bar{\mathbf{U}}$, $\forall \mathbf{v_h} \in \mathbf{V_h}$. then, probably, the function $\mathbb{G}(\mathrm{T}\mathbf{u_h})$ is much closer to $\mathrm{T}\mathbf{u}$ than $\mathrm{T}\mathbf{u_h}$.



These arguments form the basis of various **post-processing algorithms** that change a computed solution in accordance with some a priori knowledge of properties of the exact solution. If the error caused by violations of a priori regularity properties is dominant and the post-processing operator $\mathbb{G}$ is properly constructed, then

$$\|\mathbb{G}\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u}\| << \|\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u}\| .$$

In this case, the explicitly computable norm $\|\mathbb{G}\mathrm{T}u_h - \mathrm{T}u_h\|$ can be used to evaluate upper and lower bounds of the error.
Indeed, assume that there is a positive number $\boldsymbol{\alpha} < 1$ such that for the mapping T the estimate

$$\|\mathbb{G}\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u}\| \leq \boldsymbol{\alpha} \|\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u}\| .$$

Two–sided estimate

Then, for $\mathbf{e} = \mathbf{u_h} - \mathbf{u}$ we have

$$
\begin{aligned}
(\mathbf{1}-\boldsymbol{\alpha})\,\|\mathrm{T}\mathbf{e}\| = (\mathbf{1}-\boldsymbol{\alpha})\,\|\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u}\| \le \\
\le \|\mathrm{T}u_h - \mathrm{T}u\| - \|\mathbb{G}\mathrm{T}u_h - \mathrm{T}u\| \le \\
\le \|\mathbf{G}\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u_h}\| \le \\
\le \|\mathbb{G}\mathrm{T}u_h - \mathrm{T}u\| + \|\mathrm{T}u_h - \mathrm{T}u\| \le \\
\le (\mathbf{1}+\boldsymbol{\alpha})\,\|\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u}\| = (\mathbf{1}+\boldsymbol{\alpha})\,\|\mathrm{T}\mathbf{e}\|\,.
\end{aligned}
$$

Thus, if $\boldsymbol{\alpha} \ll 1$, then

$$
\|\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u}\| \simeq \|\mathbb{G}\mathrm{T}\mathbf{u_h} - \mathrm{T}\mathbf{u_h}\|\,.
$$

and the right-hand can be used as an **error indicator**.

Post-processing by averaging

Post-processing operators are often constructed by averaging $\mathrm{T}\mathbf{u_h}$ on finite element patches or on the entire domain.

Integral averaging on patches

If $\mathrm{T}\mathbf{u_h} \in \mathbf{L^2}$, then post-processing operators are obtained by various averaging procedures. Let $\mathbf{\Omega_i}$ be a **patch** of $\mathbf{M_i}$ elements, i.e.,

$$\overline{\mathbf{\Omega_i}} = \bigcup \mathbf{T_{ij}}, \quad \mathbf{j = 1, 2, ...M_i}.$$

Let $\mathbf{P^k}(\mathbf{\Omega_i}, \mathbb{R}^{\mathbf{n}})$ be a subspace of $\overline{\mathbf{U}}$ that consists of vector-valued polynomial functions of degrees less than or equal to $\mathbf{k}$. Define $\mathbf{g_i} \in \mathbf{P^k}(\mathbf{\Omega_i}, \mathbb{R}^{\mathbf{n}})$ as the minimizer of the problem:

$$\inf_{\mathbf{g} \in \mathbf{P^k}(\mathbf{\Omega_i}, \mathbb{R}^{\mathbf{n}})} \int_{\mathbf{\Omega_i}} |\mathbf{g} - \mathrm{T}\mathbf{u_h}|^2 \, \mathbf{dx}.$$

The minimizer $\mathbf{g_i}$ is used to define the values of an averaged function at some points (nodes). Further, these values are utilized by a prolongation procedure that defines an averaged function

$$\mathbb{G}\mathrm{T}\mathbf{u_h} : \mathbf{\Omega} \to \mathbb{R}.$$

Consider the simplest case. Let T be the operator ∇ and $\mathbf{u_h}$ be a piecewise affine continuous function. Then,

$$\nabla \mathbf{u_h} \in \mathbf{P^0}(\mathbf{T_{ij}}, \mathbb{R}^{\mathbf{n}}) \quad \text{on each } \mathbf{T_{ij}} \subset \mathbf{\Omega_i}.$$

We denote the values of $\nabla \mathbf{u_h}$ on $\mathbf{T_{ij}}$ by $(\nabla \mathbf{u_h})_{\mathbf{ij}}$.

Set $\mathbf{k = 0}$ and find $\mathbf{g_i} \in \mathbf{P^0}$ such that

$$\int_{\Omega_i} |\mathbf{g_i} - \nabla \mathbf{u_h}|^2 \, \mathbf{dx} = \inf_{\mathbf{g} \in \mathbf{P^0}(\Omega_i)} \int_{\Omega_i} |\mathbf{g} - \nabla \mathbf{u_h}|^2 \, \mathbf{dx} =$$

$$= \inf_{\mathbf{g} \in \mathbf{P^0}(\Omega_i)} \left\{ |\mathbf{g}|^2 |\Omega_i| - 2\mathbf{g} \cdot \sum_{j=1}^{M_i} (\nabla \mathbf{u_h})_{ij} |\mathbf{T_{ij}}| + \sum_{j=1}^{M_i} |(\nabla \mathbf{u_h})_{ij}|^2 |\mathbf{T_{ij}}| \right\}.$$

It is easy to see that $\mathbf{g_i}$ is given by a weighted sum of $(\nabla \mathbf{u_h})_{ij}$, namely,

$$\mathbf{g_i} = \sum_{j=1}^{M_i} \frac{|\mathbf{T_{ij}}|}{|\Omega_i|} (\nabla \mathbf{u_h})_{ij}.$$

Set $\mathbb{G}(\nabla \mathbf{u_h})(\mathbf{x_i}) = \mathbf{g_i}$.

Repeat this procedure for all nodes and define the vector-valued function $\mathbb{G}\nabla(\mathbf{u_h})$ by the piecewise affine prolongation of these values. For regular meshes with equal $|\mathbf{T_{ij}}|$, we have

$$\mathbf{g_i} = \sum_{\mathbf{j=1}}^{\mathbf{M_i}} \frac{\mathbf{1}}{\mathbf{M_i}} (\nabla \mathbf{u_h})_{\mathbf{ij}}.$$

Various averaging formulas of this type are represented in the form

$$\mathbf{g_i} = \sum_{\mathbf{j=1}}^{\mathbf{M_i}} \lambda_{\mathbf{ij}} (\nabla \mathbf{u_h})_{\mathbf{ij}}, \quad \sum_{\mathbf{j=1}}^{\mathbf{M_i}} \lambda_{\mathbf{ij}} = \mathbf{1},$$

where $\lambda_{\mathbf{ij}}$ are the weight factors. For internal nodes, they may be taken, e.g., as follows

$$\lambda_{\mathbf{ij}} = \frac{|\gamma_{\mathbf{ij}}|}{\mathbf{2}\pi}, \quad |\gamma_{\mathbf{ij}}| \text{ is the angle.}$$

However, if a node **belongs to the boundary**, then it is better to choose **special weights**. Their values depend on the mesh and on the type of the boundary. Concerning this point see

I. Hlavácek and M. Krizek. On a superconvergence finite element scheme for elliptic systems. I. Dirichlet boundary conditions. *Aplikace Matematiky*, 32(1987), No.2, 131-154.

Discrete averaging on patches

Consider the problem

$$\inf_{g \in \mathbb{P}^k(\Omega_i)} \sum_{s=1}^{m_i} |g(x_s) - T u_h(x_s)|^2 ,$$

where the points x_s are specially selected in Ω_i. Usually, the points x_s are the so–called **superconvergent points**.
Let $g_i \in \mathbb{P}^k(\Omega_i)$ be the minimizer of this problem.
If $k = 0$, and $T = \nabla$ then

$$g_i = \frac{1}{m_i} \sum_{s=1}^{m_i} \nabla u_h(x_s).$$

Global averaging

Global averaging makes the post–processing not on patches, but on the whole domain.

Assume that $\mathrm{T}\mathbf{u_h} \in \mathbf{L^2}$ and find $\bar{\mathbf{g}}_\mathbf{h} \in \mathbf{V_h}(\Omega) \subset \overline{\mathbf{U}}$ such that

$$\|\bar{\mathbf{g}}_\mathbf{h} - \mathrm{T}\mathbf{u_h}\|^2_\Omega = \inf_{\mathbf{g_h} \in \mathbf{V_h}(\Omega)} \|\mathbf{g_h} - \mathrm{T}\mathbf{u_h}\|^2_\Omega \, .$$

The function $\bar{\mathbf{g}}_\mathbf{h}$ can be viewed as $\mathbb{G}\mathrm{T}\mathbf{u_h}$. Very often $\bar{\mathbf{g}}_\mathbf{h}$ is a better image of $\mathrm{T}\mathbf{u}$ than the functions obtained by local procedures.

Remark

Moreover, mathematical justifications of the methods based on global averaging procedures can be performed under weaker assumptions what makes them applicable to a wider class of problems see, e.g.,

Carstensen, C.; Bartels, S. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. I: Low order conforming, nonconforming, and mixed FEM, *Math. Comp.*, 71(2002)

Task 5

Solve the boundary–value problem

$$\mathbf{\Delta u} + \mathbf{f} = \mathbf{0}, \quad \mathbf{u} = \mathbf{0} \text{ on } \partial\mathbf{\Omega}$$

by *h*-version FEM (use Matlab or another code). Apply the simplest gradient–averaging error indicator to indicate the error distribution. Compare it with the distribution of true error (the latter can be extracted from a solution on a much finer mesh).

Justifications of the method. Superconvergence

Let $\mathbf{u_h}$ be a Galerkin approximation of $\mathbf{u}$ computed on $\mathbf{V_h}$. For piecewise affine approximations of the diffusion problem we have the estimate

$$\|\nabla(\mathbf{u}-\mathbf{u_h})\|_{2,\Omega} \le \mathbf{c_1 h}, \quad \|\mathbf{u}-\mathbf{u_h}\|_{2,\Omega} \le \mathbf{c_2 h^2}$$

However, it was discovered see, e.g.,
L. A. Oganesjan and L. A. Ruchovec. *Z. Vyvcisl. Mat. i Mat. Fiz.*,9(1969);
M. Zlámal. Lecture Notes. Springer, 1977;
L. B. Wahlbin. Lecture Notes. Springer, 1969 that in certain cases **this rate may be higher**. For example it may happen that

$$|\,\mathbf{u}(\mathbf{x_s})-\mathbf{u_h}(\mathbf{x_s})\,| \le \mathbf{C h^{2+\sigma}} \qquad \sigma > 0$$

at a **superconvergent point** x_s.

Certainly, existence and location of superconvergent points strongly depends on the structure of $\mathcal{T}_h$.

For the paradigm of the diffusion problem we say that an operator $\mathbb{G}$ possesses a *superconvergence* property in $\omega \subset \Omega$ if

$$\|\nabla \mathbf{u} - \mathbb{G}\nabla \mathbf{u_h}\|_{2,\omega} \leq \mathbf{c_2 h^{1+\sigma}},$$

where the constant $\mathbf{c_2}$ may depend on higher norms of $\mathbf{u}$ and the structure of $\mathcal{T}_\mathbf{h}$.

For the diffusion problem estimates of such a type can be found, e.g., in

I. Hlavácek and M. Krizek. On a superconvergence finite element scheme for elliptic systems. I. Dirichlet boundary conditions. *Aplikace Matematiky*, 32(1987).

M. Krízek and P. Neittaanmäki. Superconvergence phenomenon in the finite element method arising from averaging of gradients *Numer. Math.*, 45(1984)

By exploiting the superconvergence properties, e.g.,

$$\|\nabla \mathbf{u} - \mathbb{G}\nabla \mathbf{u_h}\|_{2,\omega} \leq \mathbf{c_2 h^{1+\sigma}},$$

while

$$\|\nabla \mathbf{u} - \nabla \mathbf{u_h}\|_{2,\omega} \leq \mathbf{c_2 h},$$

one can usually construct a simple post-processing operator $\mathbb{G}$ satisfying the condition

$$\|\mathbb{G}\nabla \mathbf{u_h} - \nabla \mathbf{u}\| \leq \boldsymbol{\alpha} \|\nabla \mathbf{u_h} - \nabla \mathbf{u}\|.$$

where the value of $\boldsymbol{\alpha}$ decreases as $\mathbf{h}$ tends to zero.

Since

$$\|\mathbb{G}\nabla \mathbf{u_h} - \nabla \mathbf{u_h}\| \leq \|\nabla \mathbf{u_h} - \nabla \mathbf{u}\| + \|\mathbb{G}\nabla \mathbf{u_h} - \nabla \mathbf{u}\|,$$
$$\|\mathbb{G}\nabla \mathbf{u_h} - \nabla \mathbf{u_h}\| \geq \|\nabla \mathbf{u_h} - \nabla \mathbf{u}\| - \|\mathbb{G}\nabla \mathbf{u_h} - \nabla \mathbf{u}\|.$$

where the first term in the right–hand side is of the order $\mathbf{h}$ and the second one is of $\mathbf{h}^{1+\delta}$. We see that

$$\|\mathbb{G}\nabla \mathbf{u_h} - \nabla \mathbf{u_h}\| \sim \mathbf{h}$$

Therefore, we observe that in the decomposition

$$\|\nabla(\mathbf{u_h} - \mathbf{u})\| \leq \|\nabla \mathbf{u_h} - \mathbb{G}\nabla \mathbf{u_h}\| + \|\mathbb{G}\nabla \mathbf{u_h} - \nabla \mathbf{u}\|$$

asymptotically dominates the second directly computable term.

Thus, we obtain a simple error indicator:

$$\|\nabla(\mathbf{u_h} - \mathbf{u})\| \approx \|\nabla \mathbf{u_h} - \mathbb{G}\nabla \mathbf{u_h}\| .$$

Note that

$$\mathbf{i_{eff}} = \frac{\|\nabla(\mathbf{u_h} - \mathbf{u})\|}{\|\nabla \mathbf{u_h} - \mathbb{G}\nabla \mathbf{u_h}\|} \approx \mathbf{1} + \mathbf{ch}^{\delta}$$

so that this error indicator is **asymptotically exact** provided that $\mathbf{u_h}$ is a Galerkin approximation, $\mathbf{u}$ is sufficiently regular and $\mathbf{h}$ is small enough. Such type error indicators (often called **ZZ indicators** by the names of Zienkiewicz and Zhu) are widely used as cheap error indicators in engineering computations.

Some references

M. Ainsworth, J. Z. Zhu, A. W. Craig and O. C. Zienkiewicz. Analysis of the Zienkiewicz-Zhu a posteriori error estimator in the finite element method, *Int. J. Numer. Methods Engrg.*, 28(1989).

I. Babuska and R. Rodriguez. The problem of the selection of an a posteriori error indicator based on smoothing techniques, *Internat. J. Numer. Meth. Engrg.*, 36(1993).

O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis, *Internat. J. Numer. Meth. Engrg.*, 24(1987)

Post-processing by equilibration

For a solution of the diffusion problem we know that

$$\mathbf{div}\boldsymbol{\sigma} + \mathbf{f} = \mathbf{0},$$

where $\boldsymbol{\sigma} = \mathbf{A}\nabla\mathbf{u}$. This suggests an idea to construct an operator $\mathbb{G}$ such that

$$\mathbf{div}(\mathbb{G}(\mathbf{A}\nabla\mathbf{u_h})) + \mathbf{f} = \mathbf{0}.$$

If $\mathbb{G}$ possesses additional properties (linearity, boundedness), then we may hope that the function $\mathbb{G}\mathbf{A}\nabla\mathbf{u_h}$ is closer to $\boldsymbol{\sigma}$ than $\mathbf{A}\nabla\mathbf{u_h}$ and use the quantity $\|\mathbf{A}\nabla\mathbf{u_h} - \mathbb{G}\mathbf{A}\nabla\mathbf{u_h}\|$ as an error indicator.

This idea can be applied to an important class of problems

$$\mathbf{\Lambda}^{\star}\mathrm{T}u + f = 0, \qquad \mathrm{Tu} = \mathcal{A}\mathbf{\Lambda}u, \tag{56}$$

where $\mathcal{A}$ is a positive definite operator, $\mathbf{\Lambda}$ is a linear continuous operator, and $\mathbf{\Lambda}^{\star}$ is the adjoint operator.
In continuum mechanics, equations of the type (56) are referred to as the **equilibrium equations**. Therefore, it is natural to call an operator $\mathbb{G}$ an **equilibration** operator.

If the equilibration has been performed exactly then it is not difficult to get an upper error bound. However, in general, this task is either cannot be fulfilled or lead to complicated and expensive procedures. Known methods are usually end with approximately equilibrated fluxes.

Goal–oriented error estimates

Global error estimates give a general idea on the quality of an approximate solution and stopping criteria. However, often it is useful to estimate the errors in terms of **specially selected linear functionals** ℓ_s, $s = 1, 2, ...M$, e.g.,

$$< \ell, \mathbf{v} - \mathbf{u} >= \int_{\Omega} \varphi_0 \left(\mathbf{v} - \mathbf{u}\right) \mathbf{dx},$$

where ϕ is a locally supported function. Since

$$| < \ell, \mathbf{u} - \mathbf{u_h} > | \leq \|\ell\| \|\mathbf{u} - \mathbf{u_h}\|_{\mathbf{V}},$$

we can obtain such an estimate throughout the global a posteriori estimate. However, in many cases, such a method will strongly overestimate the quantity.

Adjoint problem

A posteriori estimates of the errors evaluated in terms of linear functionals are derived by attracting the **adjoint** boundary-value problem whose right-hand side is formed by the functional ℓ.
Let us represent this idea in the simplest form. Consider a system

$$\mathbf{A}\mathbf{u} = \mathbf{f},$$

where $\mathbf{A}$ is a positive definite matrix and $\mathbf{f}$ is a given vector. Let $\mathbf{v}$ be an approximate solution. Define $\mathbf{u}_\ell$ by the relation

$$\mathbf{A}^\star \mathbf{u}_\ell = \ell,$$

where $\mathbf{A}^\star$ is the matrix adjoint to $\mathbf{A}$. Then,

$$\ell \cdot (\mathbf{u} - \mathbf{v}) = \mathbf{A}^\star \mathbf{u}_\ell \cdot \mathbf{u} - \ell \cdot \mathbf{v} = \mathbf{f} \cdot \mathbf{u}_\ell - \ell \cdot \mathbf{v} = (\mathbf{f} - \mathbf{A}\mathbf{v}) \cdot \mathbf{u}_\ell$$

Certainly, the above consideration holds in a more general (operator) sense, so that for a pair of operators A and $\mathbf{A}^\star$ we have

$$< \ell, \mathbf{u} - \mathbf{v} > = < \mathbf{f} - \mathbf{A}\mathbf{v}, \mathbf{u}_\ell > . \tag{57}$$

and find the error with respect to a linear functional by the product of the residual and the exact solution of the adjoint problem:

$$\mathbf{A}^\star \mathbf{u}_\ell = \ell.$$

Practical application of this principle depends on the ability to find either $\mathbf{u}_\ell$ or its sharp approximation.

Consider again the diffusion problem. Now, it is convenient to denote the solution of the original problem by $\mathbf{u_f}$, i.e

$$\int_\Omega \mathbf{A}\nabla \mathbf{u_f} \cdot \nabla \mathbf{w}\, \mathbf{dx} = \int_\Omega \mathbf{f w}\, \mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0}(\mathbf{\Omega}).$$

Since in our case $\mathbf{A} = \mathbf{A}^\star$, the **adjoint** problem is to find $\mathbf{u}_\ell \in \mathbf{V_0}(\mathbf{\Omega})$ such that

$$\int_\Omega \mathbf{A}\nabla \mathbf{u}_\ell \cdot \nabla \mathbf{w}\, \mathbf{dx} = \int_\Omega \ell \mathbf{w}\, \mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0}(\mathbf{\Omega}).$$

Let Ω be divided into a number of elements $\mathbf{T_i}$, $\mathbf{i = 1, 2, ...N}$. Given approximations on the elements, we define a finite-dimensional subspace $\mathbf{V_{0h} \in V_0(\Omega)}$ and the Galerkin approximations $\mathbf{u_{fh}}$ and $\mathbf{u_{\ell h}}$:

$$\int_\Omega \mathbf{A}\nabla \mathbf{u_{fh}} \cdot \nabla \mathbf{w_h}\, \mathbf{dx} = \int_\Omega \mathbf{f w_h dx}, \qquad \forall \mathbf{w_h} \in \mathbf{V_{0h}},$$

$$\int_\Omega \mathbf{A}\nabla \mathbf{u_{\ell h}} \cdot \nabla \mathbf{w_h}\, \mathbf{dx} = \int_\Omega \ell \mathbf{w_h dx}, \qquad \forall \mathbf{w_h} \in \mathbf{V_{0h}}.$$

Since

$$\int_\Omega \ell(\mathbf{u_f} - \mathbf{u_{fh}})\mathbf{dx} = \int_\Omega \mathbf{A}\nabla \mathbf{u}_\ell \cdot \nabla(\mathbf{u_f} - \mathbf{u_{fh}})\mathbf{dx}$$

and

$$\int_\Omega \mathbf{A}\nabla \mathbf{u_{\ell h}} \cdot \nabla(\mathbf{u_f} - \mathbf{u_{fh}})\mathbf{dx} = \mathbf{0},$$

We arrive at the relation

$$\int_\Omega \ell(\mathbf{u_f} - \mathbf{u_{fh}})\mathbf{dx} = \int_\Omega \mathbf{A}\nabla(\mathbf{u}_\ell - \mathbf{u}_{\ell\mathbf{h}}) \cdot \nabla(\mathbf{u_f} - \mathbf{u_{fh}})\mathbf{dx} \qquad (58)$$

whose right-hand side is expressed in the form

$$\sum_{\mathbf{i}=1}^{\mathbf{N}} \int_{\mathbf{T_i}} \mathbf{A}\nabla(\mathbf{u_f} - \mathbf{u_{fh}}) \cdot \nabla(\mathbf{u}_\ell - \mathbf{u}_{\ell\mathbf{h}})\,\mathbf{dx} =$$

$$\sum_{\mathbf{i}=1}^{\mathbf{N}} \left\{ -\int_{\mathbf{T_i}} \mathbf{div}\left(\mathbf{A}\nabla(\mathbf{u_f} - \mathbf{u_{fh}})\right)(\mathbf{u}_\ell - \mathbf{u}_{\ell\mathbf{h}})\,\mathbf{dx} + \right.$$

$$\left. + \frac{\mathbf{1}}{\mathbf{2}} \int_{\partial \mathbf{T_i}} \mathbf{j}\left(\nu_\mathbf{i} \cdot \mathbf{A}\nabla(\mathbf{u_f} - \mathbf{u_{fh}})\right)(\mathbf{u}_\ell - \mathbf{u}_{\ell\mathbf{h}})\,\mathbf{ds} \right\}.$$

This relation implies the estimate

$$\int_\Omega \ell(\mathbf{u_f}-\mathbf{u_{fh}})\mathbf{dx}=\sum_{\mathbf{i}=1}^{\mathbf{N}}\Big\{\|\mathbf{divA}\nabla(\mathbf{u_f}-\mathbf{u_{fh}})\|_{\mathbf{2,T_i}}\,\|\mathbf{u}_\ell-\mathbf{u}_{\ell\mathbf{h}}\|_{\mathbf{2,T_i}}+$$

$$+\tfrac{\mathbf{1}}{\mathbf{2}}\,\|\mathbf{j}(\boldsymbol{\nu}_\mathbf{i}\cdot\mathbf{A}\nabla(\mathbf{u_f}-\mathbf{u_{fh}}))\|_{\mathbf{2},\partial\mathbf{T_i}}\;\|\mathbf{u}_\ell-\mathbf{u}_{\ell\mathbf{h}}\|_{\mathbf{2},\partial\mathbf{T_i}}\Big\}=$$

$$=\sum_{\mathbf{i}=1}^{\mathbf{N}}\Big\{\|\mathbf{f}+\mathbf{divA}\nabla\mathbf{u_{fh}}\|_{\mathbf{2,T_i}}\;\|\mathbf{u}_\ell-\mathbf{u}_{\ell\mathbf{h}}\|_{\mathbf{2,T_i}}\;+$$

$$+\tfrac{\mathbf{1}}{\mathbf{2}}\,\|\mathbf{j}(\boldsymbol{\nu}_\mathbf{i}\cdot\mathbf{A}\nabla\mathbf{u_{fh}})\|_{\mathbf{2},\partial\mathbf{T_i}}\;\|\mathbf{u}_\ell-\mathbf{u}_{\ell\mathbf{h}}\|_{\mathbf{2},\partial\mathbf{T_i}}\Big\}.$$

Here, the principal terms are the same as in the explicit residual method, but the weights are given by the norms of $\mathbf{u}_\ell-\mathbf{u}_{\ell\mathbf{h}}$.

Assume that $\mathbf{u}_\ell \in \mathbf{H}^2$ and $\mathbf{u}_{\ell\mathbf{h}}$ is constructed by piecewise affine continuous approximations. Then the norms $\|\mathbf{u}_\ell - \mathbf{u}_{\ell\mathbf{h}}\|_{\mathbf{T_i}}$ and $\|\mathbf{u}_\ell - \mathbf{u}_{\ell\mathbf{h}}\|_{\mathbf{2},\partial\mathbf{T_i}}$ are estimated by the quantities $h^{\alpha}|u_\ell|_{2,2,T_i}$ with $\alpha = 1$ and $1/2$ and the multipliers $\hat{\mathbf{c}}_\mathbf{i}$ and $\hat{\mathbf{c}}_\mathbf{ij}$, respectively.
In this case, we obtain an estimate with constants defined by the standard

$$\mathbf{H}^2 \longrightarrow \mathbf{V}_{\mathbf{0h}}$$

interpolation operator whose evaluation is much simpler than that of the constants arising in the

$$\mathbf{H}^1 \longrightarrow \mathbf{V}_{\mathbf{0h}}$$

interpolation.

A posteriori estimates in L^2–norm

In principle, this technology can be exploited to evaluate estimates in $\mathbf{L^2}$–norm. Indeed,

$$\|\mathbf{u_f} - \mathbf{u_{fh}}\| = \sup_{\ell \in \mathbf{L^2}} \frac{(\ell, \mathbf{u_f} - \mathbf{u_{fh}})}{\|\ell\|} = \sup_{\ell \in \mathbf{L^2}} \frac{(\mathbf{A}\nabla \mathbf{u}_\ell, \nabla(\mathbf{u_f} - \mathbf{u_{fh}}))}{\|\ell\|} =$$

$$= \sup_{\ell \in \mathbf{L^2}} \frac{(\mathbf{A}\nabla(\mathbf{u}_\ell - \pi_\mathbf{h}(\mathbf{u}_\ell)), \nabla(\mathbf{u_f} - \mathbf{u_{fh}}))}{\|\ell\|} =$$

$$= \sup_{\ell \in \mathbf{L^2}} \frac{(\nabla(\mathbf{u}_\ell - \pi_\mathbf{h}(\mathbf{u}_\ell)), \mathbf{A}\nabla(\mathbf{u_f} - \mathbf{u_{fh}}))}{\|\ell\|} =$$

$$= \sup_{\ell \in \mathbf{L^2}} \frac{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{N}} \left\{ \int_{\mathbf{T_i}} \nabla(\mathbf{u}_\ell - \pi_\mathbf{h}(\mathbf{u}_\ell)), \mathbf{A}\nabla(\mathbf{u_f} - \mathbf{u_{fh}})\, \mathbf{dx} \right\}}{\|\ell\|}$$

Integrating by parts, we obtain

$$\frac{\sum_{i=1}^{N} \left\{ \|\mathbf{f}+\mathbf{div}\mathbf{A}\nabla\mathbf{u_{fh}}\|_{\mathbf{T_i}} \ \|\mathbf{u}_\ell-\pi_\mathbf{h}(\mathbf{u}_\ell)\|_{\mathbf{T_i}} + \frac{\mathbf{1}}{\mathbf{2}} \|\mathbf{j}(\nu_\mathbf{i}\cdot\mathbf{A}\nabla\mathbf{u_{fh}})\|_{\partial\mathbf{T_i}} \|\mathbf{u}_\ell-\pi_\mathbf{h}(\mathbf{u}_\ell)\|_{\partial\mathbf{T_i}} \right\}}{\|\ell\|}$$

If for *any* $\ell \in L^2$ the adjoint problem has a regular solution (e.g., $\mathbf{u}_\ell \in \mathbf{H^2}$), so that we could combine the standard interpolation estimate for the interpolant of $\mathbf{u}_\ell$ with the regularity estimate for the PDE (e.g., $\|\mathbf{u}_\ell\| \le \mathbf{C_1}\|\ell\|$), then we obtain

$$\|\mathbf{u}_\ell - \pi_\mathbf{h}(\mathbf{u}_\ell)\|_{\mathbf{T_i}} \le \mathbf{C_1 h}^{\alpha_1}\|\ell\|, \qquad \|\mathbf{u}_\ell - \pi_\mathbf{h}(\mathbf{u}_\ell)\|_{\partial\mathbf{T_i}} \le \mathbf{C_1 h}^{\alpha_2}\|\ell\|$$

with certain $\alpha_\mathbf{k}$.
Under the above conditions $\|\ell\|$ is reduced and we arrive at the estimate, in which the element residuals and interelement jumps are weighted with factors $\mathbf{C_1 h}^{\alpha_1}$ and $\mathbf{C_2 h}^{\alpha_2}$.

References

Methods using adjoint problems has been investigated in the works of R. Becker, C. Johnson, R. Rannacher and other scientists. A more detailed exposition of these works can be found in

W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations.* Birkhäuser, Berlin, 2003.

R. Becker and R. Rannacher. A feed–back approach to error control in finite element methods: Basic approach and examples, *East–West J. Numer. Math.*, 4(1996), 237-264.

Concerning error estimation in goal–oriented quantities we refer, e.g., to

J. T. Oden, S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element method, *Comput. Math. Appl.*, 41, 735-756, 2001.

S. Korotov, P. Neittaanmaki and S. Repin. A posteriori error estimation of goal-oriented quantities by the superconvergence patch recovery, *J. Numer. Math.* 11 (2003)

Comment

In the literature devoted to a posteriori error analysis one can find often find terms like
"duality approach to a posteriori error estimation" or
"dual-based error estimates".
However, the essence behind such a terminology may be quite different because the word *"duality"* is used in 3 different meanings:
(a) Duality in the sense of functional spaces. We have seen that if for the equation $\mathcal{L}\mathbf{u} = \mathbf{f}$ errors are measured in the original (energy) norm then a consistent upper bound is given by the residual in the norm of the space **topologically dual** to a subspace of the energy space (e.g., $\mathbf{H}^{-1}$).
(b) Duality in the sense of using the Adjoint Problem.
(c) Duality in the sense of the Theory of the Calculus of Variations.

In the next lecture
we will proceed to the detailed exposition
of the approach (c).

Lecture 3

In the lecture, we derive Functional A Posteriori Estimate for the problem

$$\mathbf{\Delta u} + \mathbf{f} = \mathbf{0},\ \Omega \qquad \mathbf{u} = \mathbf{0}\ \partial\Omega.$$

and discuss its meaning, principal features and practical implementation.

Lecture plan

1. **Functional a posteriori estimates**.
2. **How to derive them? Paradigm of a simple elliptic problem**
3. **How to use them in practice?**
4. **Examples.**

Lecture plan

1. Functional a posteriori estimates.
2. How to derive them? Paradigm of a simple elliptic problem
3. How to use them in practice?
4. Examples.

Lecture plan

1. Functional a posteriori estimates.
2. How to derive them? Paradigm of a simple elliptic problem
3. How to use them in practice?
4. Examples.

Lecture plan

1. Functional a posteriori estimates.
2. How to derive them? Paradigm of a simple elliptic problem
3. How to use them in practice?
4. Examples.

Functional A Posteriori Estimates

Functional A Posteriori Estimate is a computable majorant of the difference between exact solution $\mathbf{u}$ and any conforming approximation $\mathbf{v}$ having the general form:

$$\mathbf{\Phi}(\mathbf{u}-\mathbf{v}) \leq \mathbf{M}\left(\mathcal{D}, \mathbf{v}\right) \qquad \forall \mathbf{v} \in \mathbf{V}\,! \tag{59}$$

$\mathcal{D}$ is the data set (coefficients, domain, parameters, etc.),
$\mathbf{\Phi} : \mathbf{V} \to \mathrm{R}_+$ is a given functional.
M must be computable and continuous in the sense that

$$\mathbf{M}\left(\mathcal{D}, \mathbf{v}\right) \to \mathbf{0}, \quad \text{if } \mathbf{v} \to \mathbf{u}$$

Types of Φ

- **Energy norm** $\Phi(\mathbf{u}-\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|_{\Omega}$
- **Local norm** $\Phi(\mathbf{u}-\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|_{\omega}$
- **Goal–oriented quantity** $\Phi(\mathbf{u}-\mathbf{v}) = (\ell, \mathbf{u}-\mathbf{v})$

Functional a posteriori estimate gives complete solution of the error control problem from the viewpoint of the MATHEMATICAL THEORY of PDE's

METHODS OF THE DERIVATION.

These estimates are derived by purely functional methods using the analysis of variational problems or integral identities.

Variational method 96'-97'

Exploits variational structure of the original problem and Duality Theory in the Calculus of Variations.

See

S. Repin *Mathematics of Computation*, 69(230), pp. 2000, 481-500.

A systematic exposition of the variational approach to deriving Functional a Posteriori Estimates can be found in

P. Neittaanmaki and S. Repin. *Reliable methods for computer simulation. Error control and a posteriori estimates.* Elsevier, NY, 2004

Nonvariational method 2000'

Derives a posteriori estimates by certain transformations of integral identities.

Basic idea of the method is presented in

S. Repin. *Proc. St.-Petersburg Math. Society*, 2001 pp. 148-179 (in Russian, translated in American Mathematical Translations Series 2, 9(2003)). .

Other publications:

S. Repin. Estimates of deviation from exact solutions of initial-boundary value problems for the heat equation, *Rend. Mat. Acc. Lincei*, 13(2002), pp. 121-133.

S. Repin *Estimates of deviations from exact solutions for some boundary–value problems with incompressibility condition, Algebra and Analiz*, 16(2004), 5, pp. 124–161.

Let us consider both methods in application to our basic problem

Variational Method

Let $\mathbf{u}$ be a (generalized) solution of the problem

$$\mathbf{\Delta u + f = 0},\ \mathbf{\Omega} \qquad \mathbf{u = 0}\ \partial\mathbf{\Omega}.$$

As we have seen in Lecture 1, this problem is equivalent to the following variational problem:

Problem $\mathcal{P}$. Find $\mathbf{u} \in \mathbf{V_0} := \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\mathbf{\Omega})$ such that

$$\mathbf{J(u)} = \inf_{\mathbf{v} \in \mathbf{V_0}} \mathbf{J(v)},$$

where

$$\mathbf{J(v)} = \frac{\mathbf{1}}{\mathbf{2}} \|\nabla \mathbf{v}\|^2 - (\mathbf{f}, \mathbf{v}).$$

By the reasons that we discussed earlier this problem has a unique solution.

Lagrangian

Note that

$$\mathbf{J}(\mathbf{v}) = \sup_{\mathbf{y}\in\mathbf{Y}} \mathbf{L}(\nabla\mathbf{v},\mathbf{y}), \quad \mathbf{L}(\nabla\mathbf{v},\mathbf{y}) = \int_{\Omega}\left(\nabla\mathbf{v}\cdot\mathbf{y} - \frac{1}{2}|\mathbf{y}|^2 - \mathbf{f}\mathbf{v}\right)\mathbf{dx}$$

where $\mathbf{Y} = \mathbf{L}^2(\Omega,\mathbb{R}^n)$. Indeed, the value of the above supremum cannot exceed the one we obtain if for almost all $\mathbf{x}\in\Omega$ solve the pointwise problems

$$\sup_{\mathbf{y}(\mathbf{x})} (\nabla\mathbf{v})(\mathbf{x})\cdot\mathbf{y}(\mathbf{x}) - \frac{1}{2}|\mathbf{y}(\mathbf{x})|^2 \qquad \mathbf{x}\in\Omega$$

whose upper bound is attained if set $\mathbf{y}(\mathbf{x}) = (\nabla\mathbf{v})(\mathbf{x})$. Since $\nabla\mathbf{v}\in\mathbf{Y}$, we observe that the respective maximizer belongs to $\mathbf{Y}$ and, therefore

$$\sup_{\mathbf{y}\in\mathbf{Y}} \mathbf{L}(\nabla\mathbf{v},\mathbf{y}) = \mathbf{L}(\nabla\mathbf{v},\nabla\mathbf{v}) = \mathbf{J}(\mathbf{v}).$$

Minimax Formulations

Then, the original problem comes in the minimax form:

$$(\mathcal{P}) \qquad \inf_{\mathbf{v}\in\mathbf{V_0}} \sup_{\mathbf{y}\in\mathbf{Y}} \mathbf{L}(\nabla\mathbf{v},\mathbf{y})$$

If the order of **inf** and **sup** is changed, then we arrive at the so-called **dual problem**

$$(\mathcal{P}^*) \qquad \sup_{\mathbf{y}\in\mathbf{Y}} \inf_{\mathbf{v}\in\mathbf{V_0}} \mathbf{L}(\nabla\mathbf{v},\mathbf{y})$$

Note that

$$\inf_{\mathbf{v}\in\mathbf{V_0}} \int_\Omega \left(\nabla\mathbf{v}\cdot\mathbf{y} - \frac{\mathbf{1}}{\mathbf{2}}|\mathbf{y}|^2 - \mathbf{fv}\right)\mathbf{dx} = -\frac{\mathbf{1}}{\mathbf{2}}\|\mathbf{y}\|^2 + \inf_{\mathbf{v}\in\mathbf{V_0}} \int_\Omega (\nabla\mathbf{v}\cdot\mathbf{y} - \mathbf{fv})\mathbf{dx} =$$

$$= \begin{cases} -\frac{1}{2}\|\mathbf{y}\|^2 & \text{if } \mathbf{y}\in\mathbf{Q_f} := \{\mathbf{y}\in\mathbf{Y} \mid \mathbf{divy} + \mathbf{f} = \mathbf{0}\} \\ -\infty & \text{if } \mathbf{y}\notin\mathbf{Q_f} \end{cases}$$

Dual Problem

Thus, we observe that the dual problem has the form: find $\mathbf{p} \in \mathbf{Q_f}$ such that

$$-\mathbf{I}^*(\mathbf{p}) = \sup_{\mathbf{y} \in \mathbf{Q_f}} -\mathbf{I}^*(\mathbf{y})$$

where

$$\mathbf{I}^*(\mathbf{q}) = \frac{1}{2}\|\mathbf{q}\|^2$$

How are these two problems related?

First, we establish one relation that holds regardless of the structure of the Lagrangian.

Sup Inf and Inf Sup

Lemma

Let $\mathbf{L}(\mathbf{x},\mathbf{y})$ *be a functional defined on the elements of two nonempty sets* $\mathbf{X}$ *and* $\mathbf{Y}$. *Then*

$$\sup_{\mathbf{y}\in\mathbf{Y}}\inf_{\mathbf{x}\in\mathbf{X}}\mathbf{L}(\mathbf{x},\mathbf{y}) \le \inf_{\mathbf{x}\in\mathbf{X}}\sup_{\mathbf{y}\in\mathbf{Y}}\mathbf{L}(\mathbf{x},\mathbf{y}). \tag{60}$$

Proof

It is easy to see that

$$\mathbf{L}(\mathbf{x},\mathbf{y}) \ge \inf_{\xi\in\mathbf{X}}\mathbf{L}(\xi,\mathbf{y}), \quad \forall \mathbf{x}\in\mathbf{X},\ \mathbf{y}\in\mathbf{Y}.$$

Taking the supremum over $\mathbf{y}\in\mathbf{Y}$, we obtain

proof

$$\sup_{y \in Y} L(x, y) \geq \sup_{y \in Y} \inf_{\xi \in X} L(\xi, y), \quad \forall x \in X.$$

The left-hand side depends on x, while the right-hand side is a number. Thus, we may take infimum over $x \in X$ and obtain the inequality

$$\inf_{x \in X} \sup_{y \in Y} L(x, y) \geq \sup_{y \in Y} \inf_{\xi \in X} L(\xi, y).$$

Therefore, we always have

$$\sup \mathcal{P}^* \leq \inf \mathcal{P}$$

Duality relations

However, in our case we have a stronger relation, namely

$$\sup \mathcal{P}^* = \inf \mathcal{P}$$

To prove this fact, we note that

$$\int_\Omega \nabla \mathbf{u} \cdot \nabla \mathbf{v} \, \mathbf{dx} = \int_\Omega \mathbf{f} \mathbf{v} \, \mathbf{dx} \quad \forall \mathbf{v} \in \mathbf{V_0}.$$

Therefore $\mathbf{p} = \nabla \mathbf{u} \in \mathbf{Q_f}$ and

$$-\mathbf{I}^*(\mathbf{p}) = -\frac{\mathbf{1}}{\mathbf{2}} \|\nabla \mathbf{u}\|^2 = \int_\Omega (\frac{\mathbf{1}}{\mathbf{2}} |\nabla \mathbf{u}|^2 - |\nabla \mathbf{u}|^2) \mathbf{dx} = \int_\Omega (\frac{\mathbf{1}}{\mathbf{2}} |\nabla \mathbf{u}|^2 - \mathbf{f} \mathbf{u}) \mathbf{dx} = \mathbf{J}(\mathbf{u}).$$

Let us use the Mikhlin's estimate established in Lecture 2:

$$\frac{1}{2}\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \le \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}).$$

Since $\mathbf{J}(\mathbf{u}) = -\mathbf{I}^*(\mathbf{p})$, we have

$$\frac{1}{2}\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \le \mathbf{J}(\mathbf{v}) + \mathbf{I}^*(\mathbf{p}) \le \mathbf{J}(\mathbf{v}) + \mathbf{I}^*(\mathbf{q}) \; \forall \mathbf{q} \in \mathbf{Q_f}.$$

Reform this estimate by using the fact that $\mathbf{q} \in \mathbf{Q_f}$.

$$\begin{aligned}\mathbf{J}(\mathbf{v}) + \mathbf{I}^*(\mathbf{q}) &= \frac{1}{2}\|\nabla\mathbf{v}\|^2 - (\mathbf{f},\mathbf{v}) + \frac{1}{2}\|\mathbf{q}\|^2 \\ &= \frac{1}{2}\|\nabla\mathbf{v}\|^2 + \frac{1}{2}\|\mathbf{q}\|^2 - (\nabla\mathbf{v},\mathbf{q}) = \\ &= \frac{1}{2}\|\nabla\mathbf{v} - \mathbf{q}\|^2.\end{aligned}$$

Now, we have

$$\|\nabla(\mathbf{v}-\mathbf{u})\| \le \|\nabla\mathbf{v}-\mathbf{q}\| \qquad \forall \mathbf{q} \in \mathbf{Q_f}.$$

Take arbitrary $\mathbf{y} \in \mathbf{L^2(\Omega)}$. Then,

$$\|\nabla(\mathbf{v}-\mathbf{u})\| \le \|\nabla\mathbf{v}-\mathbf{y}\| + \inf_{\mathbf{q}\in\mathbf{Q_f}} \|\mathbf{y}-\mathbf{q}\|.$$

How to estimate the above infimum?

Various methods give one and the same answer:

$$\inf_{\mathbf{q}\in\mathbf{Q_f}} \|\mathbf{y}-\mathbf{q}\| \le \mathbf{[}\,\mathbf{div y}+\mathbf{f}\,\mathbf{]} \qquad \mathbf{y} \in \mathbf{L^2(\Omega)}, \tag{61}$$

$$\inf_{\mathbf{q}\in\mathbf{Q_f}} \|\mathbf{y}-\mathbf{q}\| \le \mathbf{C_\Omega}\|\mathbf{div y}+\mathbf{f}\| \qquad \mathbf{y} \in \mathbf{H(\Omega, div)}, \tag{62}$$

Proof

To prove these estimates we consider an auxiliary problem

$$\mathbf{\Delta}\eta + \mathbf{f} + \mathbf{div}\mathbf{y} = \mathbf{0} \ \ \Omega \quad \eta = \mathbf{0} \ \ \partial\Omega.$$

$$\int_\Omega \nabla\eta \cdot \nabla \mathbf{w} d\mathbf{x} = \int_\Omega (\mathbf{f}\mathbf{w} - \mathbf{y} \cdot \nabla \mathbf{w}) d\mathbf{x}$$

$\overline{\mathbf{q}}$

$$\int_\Omega \overbrace{(\nabla\eta + \mathbf{y})} \cdot \nabla \mathbf{w} \, d\mathbf{x} = \int_\Omega \mathbf{f}\mathbf{w} \, d\mathbf{x} \quad \forall \mathbf{w} \in \mathbf{V_0}$$

Thus, $\bar{\mathbf{q}} \in Q_f$!!!

Since $\boldsymbol{\eta}$ is a solution of the boundary–value problem with right–hand side $\mathbf{div}\,\mathbf{y} + \mathbf{f} \in \mathbf{H}^{-1}$, we have

$$\|\nabla \boldsymbol{\eta}\| \;\leq\; \mathbf{[}\,\mathbf{div}\,\mathbf{y} + \mathbf{f}\,\mathbf{]},$$

Then

$$\inf_{\mathbf{q}\in\mathbf{Q_f}} \|\mathbf{y} - \mathbf{q}\| \leq \|\mathbf{y} - \bar{\mathbf{q}}\| = \|\nabla \boldsymbol{\eta}\| \leq \mathbf{[}\,\mathbf{divy} + \mathbf{f}\,\mathbf{]}.$$

Here

$$\mathbf{[}\,\mathbf{divy} + \mathbf{f}\,\mathbf{]} = \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\int_{\Omega}(\mathbf{y}\cdot\nabla\mathbf{w} - \mathbf{fw})\mathbf{dx}}{\|\nabla\mathbf{w}\|}$$

$\mathbf{y} \in \mathbf{H}(\Omega, \mathbf{div})$

If $\mathbf{y}$ has a square summable divergence, then we have

$$\| \mathbf{divy} + \mathbf{f} \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{divy} + \mathbf{f}) \mathbf{w} d\mathbf{x}}{\|\nabla \mathbf{w}\|} \leq \mathbf{C_\Omega} \|\mathbf{divy} + \mathbf{f}\|,$$

where $\mathbf{C_\Omega}$ is the constant in the Friederichs–Steklov inequality for the domain Ω. We observe that

a "noncomputable" negative norm has been estimated by a "computable" one without an attraction of Galerkin orthogonality and local (mesh–dependent) constants.

Thus, for any $\mathbf{y} \in \mathbf{H}(\Omega, \mathbf{div})$ we obtain

$$\|\nabla(\mathbf{v}-\mathbf{u})\| \leq \|\nabla\mathbf{v}-\mathbf{y}\| + \inf_{\mathbf{q}\in\mathbf{Q_f}} \|\mathbf{y}-\mathbf{q}\| \leq \|\nabla\mathbf{v}-\mathbf{y}\| + \mathbf{C_\Omega}\|\mathbf{div}\mathbf{y}+\mathbf{f}\|.$$

Above presented *modus operandi* can be viewed as a simplest version of the variational approach to the derivation of Functional Error Majorants.

Deriving a posteriori estimates from integral identities.

For many problems the variational techniques cannot be applied (e.g., because they may have no variational formulation).
In
S. Repin. Two-sided estimates for deviation from an exact solution to uniformly elliptic equation. *Trudi St.-Petersburg Math. Society*, 9(2001), translated in *American Mathematical Translations Series 2, 9(2003)*)
it was suggested another method, which is *based on certain transformations of integral identities.*

Non–variational method in the simplest case

Let us expose its simplest version adapted to our model problem. We have

$$\int_{\Omega} \nabla(\mathbf{u}-\mathbf{v})\nabla\mathbf{w}\mathbf{dx} = \int_{\Omega} (\mathbf{fw} - \nabla\mathbf{v}\cdot\nabla\mathbf{w})\,\mathbf{dx}$$

In order to get an upper bound of $\|\nabla(\mathbf{u}-\mathbf{v})\|$ we use the relation

$$\int_{\Omega} (\mathbf{divyw} + \nabla\mathbf{w}\cdot\mathbf{y})\,\mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{w} \in \mathbf{V_0}$$

valid for any $\mathbf{y} \in \mathbf{H}(\mathbf{\Omega}, \mathbf{div})$.

We have

$$\int_{\Omega} (\nabla \mathbf{v} \cdot \nabla \mathbf{w} - \mathbf{f}\mathbf{w}) \mathbf{dx} =$$

$$\int_{\Omega} (\nabla \mathbf{v} \cdot \nabla \mathbf{w} - \mathbf{f}\mathbf{w} - (\mathbf{divy}\, \mathbf{w} + \nabla \mathbf{w} \cdot \mathbf{y})) \mathbf{dx} =$$

$$\int_{\Omega} ((\nabla \mathbf{v} - \mathbf{y}) \cdot \nabla \mathbf{w} - (\mathbf{f} + \mathbf{divy})\mathbf{w}) \mathbf{dx} \leq$$

$$\|\nabla \mathbf{v} - \mathbf{y}\| \|\nabla \mathbf{w}\| + \|\mathbf{f} + \mathbf{divy}\| \|\mathbf{w}\| \leq$$

$$\leq (\|\nabla \mathbf{v} - \mathbf{y}\| + \mathbf{C}_{\Omega} \|\mathbf{f} + \mathbf{divy}\|) \|\nabla \mathbf{w}\|.$$

Set $\mathbf{w} = \mathbf{u} - \mathbf{v}$.

$$\int_{\Omega} |\nabla(\mathbf{u} - \mathbf{v})|^2 \mathbf{dx} \leq (\|\nabla \mathbf{v} - \mathbf{y}\| + \mathbf{C}_{\Omega} \|\mathbf{f} + \mathbf{divy}\|) \|\nabla(\mathbf{u} - \mathbf{v})\|.$$

Thus, we find that

$$\|\nabla(\mathbf{u} - \mathbf{v})\| \leq \|\nabla \mathbf{v} - \mathbf{y}\| + \mathbf{C}_{\Omega} \|\mathbf{f} + \mathbf{divy}\|.$$

Functional error estimate. Meaning and properties

For the problem

$$\mathbf{\Delta u} + \mathbf{f} = \mathbf{0}, \qquad \mathbf{u} = \mathbf{0} \,\text{on}\, \partial\mathbf{\Omega}$$

we have obtained the estimate

$$\boxed{\|\nabla(\mathbf{u} - \mathbf{v})\| \leq \|\nabla\mathbf{v} - \mathbf{y}\| + \mathbf{C_\Omega}\|\mathbf{div y} + \mathbf{f}\|} \tag{63}$$

The estimate is valid for any $\mathbf{v} \in \mathbf{V_0}$ and $\mathbf{y} \in \mathbf{H}(\mathbf{\Omega}, \mathbf{div})$
Two terms in the right–hand side have a clear sense: they **present measures of the errors in two basic relations**

$$\mathbf{p} = \nabla\mathbf{u}, \qquad \mathbf{divp} + \mathbf{f} = \mathbf{0} \quad \text{in}\, \mathbf{\Omega}$$

that jointly form the equation.

The estimate is sharp

If set $\mathbf{v} = \mathbf{0}$ and $\mathbf{y} = \mathbf{0}$, we obtain the energy estimate for the generalized solution

$$\|\nabla \mathbf{u}\| \leq \mathbf{C}_{\Omega} \|\mathbf{f}\|.$$

Therefore, no constant less than $\mathbf{C}_{\Omega}$ can be stated in the second term.
If set $\mathbf{y} = \nabla \mathbf{u}$, than the inequality holds as the equality.
Thus, we see that the estimate (63) is **sharp** in the sense that the multipliers of both terms **cannot be taken smaller** and in the set of admissible $\mathbf{y}$ there **exists a function that makes the inequality hold as equality.**

The estimate as a quadratic functional

By means of the algebraic Young's inequality

$$2ab \le \beta a^2 + \frac{1}{\beta} b^2, \qquad \beta > 0$$

we rewrite this estimate in the form

$$\|\nabla(u - v)\|^2 \le \tag{64}$$
$$\le (1+\beta)\|\nabla v - y\|^2 + \frac{1+\beta}{\beta} C_\Omega^2 \|\mathrm{div} y + f\|^2$$

For any β the right–hand side is a quadratic functional. This property makes it possible to apply well known methods for the minimization with respect to y.

Deviation Majorant

Denote the right–hand side of (64) by $\mathcal{M}_\oplus$, i.e.,

$$\mathcal{M}_\oplus(\mathbf{v},\mathbf{y},\beta,\mathbf{C}_\Omega,\mathbf{f}) := (1+\beta)\|\nabla\mathbf{v}-\mathbf{y}\|^2 + \frac{1+\beta}{\beta}\mathbf{C}_\Omega^2\|\mathbf{div}\mathbf{y}+\mathbf{f}\|^2.$$

This functional provides an upper bound for the norm of the deviation of $\mathbf{v}$ from $\mathbf{u}$. Therefore, it is natural to call it the **Deviation Majorant**.

BVP $\Delta u + f = 0$ has another variational formulation

$$\inf_{\substack{\mathbf{v}\in\mathbf{V_0},\\ \beta>0,\\ \mathbf{y}\in\mathbf{H}(\Omega,\mathrm{div}),}} \mathcal{M}_{\oplus}(\mathbf{v},\mathbf{y},\beta,\mathbf{C}_\Omega,\mathbf{f})$$

- Minimum of this functional is zero;
- it is attained if and only if $\mathbf{v} = \mathbf{u}$ and $\mathbf{y} = \mathbf{A}\nabla\mathbf{u}$!;
- $\mathcal{M}_{\oplus}$ contains only one global constant $\mathbf{C}_\Omega$, which is problem independent;

In principle, one can select certain sequences of subspaces $\{\mathbf{V_{hk}}\} \in \mathbf{V_0}$ and $\{\mathbf{Y_{hk}}\} \in \mathbf{H}(\Omega, \mathbf{div})$ and minimize the Error Majorant with respect to these subspaces

$$\inf_{\substack{\mathbf{v}\in\mathbf{V_{hk}},\\ \beta>0,\\ \mathbf{y}\in\mathbf{Y_{hk}},}} \mathcal{M}_{\oplus}(\mathbf{v}, \mathbf{y}, \beta, \mathbf{C}_{\Omega}, \mathbf{f})$$

If the subspaces are limit dense, then we would obtain a sequence of approximate solutions $(\mathbf{v_k}, \mathbf{y_k})$ and the sequence of numbers

$$\gamma_{\mathbf{k}} := \inf_{\beta>0} \mathcal{M}_{\oplus}(\mathbf{v_k}, \mathbf{y_k}, \beta, \mathbf{C}_{\Omega}, \mathbf{f}) \rightarrow \mathbf{0}$$

How to use the Majorants in practice?

Consider **CONFORMING FEM APPROXIMATIONS**.

We have 3 basic ways to use the deviation estimate:
(a) **Direct** (via flux averaging on the mesh $\mathcal{T}_h$);
(b) **One step delay** (via flux averaging on the mesh $\mathbf{h_{ref}}$);
(c) **Minimization** (minimization via $\mathbf{y}$).

(a) Use recovered gradients

Let $\mathbf{u_h} \in \mathbf{V_h}$, then

$$\mathbf{p_h} := \nabla \mathbf{u_h} \in \mathbf{L_2}(\Omega, \mathbb{R}^{\mathbf{d}}), \quad \mathbf{p_h} \notin \mathbf{H}(\Omega, \mathrm{div}).$$

Use an averaging operator $\mathbf{G_h} : \mathbf{L_2}(\Omega, \mathbb{R}^{\mathbf{d}}) \to \mathbf{H}(\Omega, \mathrm{div})$ and have a **directly computable estimate**

$$\|\nabla(\mathbf{u} - \mathbf{u_h})\| \le \|\nabla \mathbf{u_h} - \mathbf{G_h p_h}\| + \mathbf{C_\Omega} \, \|\mathbf{div G_h p_h} + \mathbf{f}\|$$

(b) Use recovered gradients from $\mathcal{T}_{h_{ref}}$

Let $\mathbf{u_1}, \mathbf{u_2}, ..., \mathbf{u_k}, ...$ be a sequence of approximations on meshes $\mathcal{T}_{h_k}$.
Compute $\mathbf{p_k} := \nabla \mathbf{u_k}$, average it by $\mathbf{G_k}$ and for $\mathbf{u_{k-1}}$ use the estimate

$$\|\mathbf{u} - \mathbf{u_{k-1}}\| \leq \|\nabla \mathbf{u_{k-1}} - \mathbf{G_k p_k}\| + \mathbf{C_\Omega} \|\mathbf{div G_k p_k} + \mathbf{f}\|$$

This estimate gives **a quantitative form of the Runge's rule.**

(c) Minimize $\mathcal{M}_\oplus$ with respect to y.

Select a certain subspace $\mathbf{Y}_\tau$ in $\mathbf{H}(\Omega, \mathbf{div})$. **Generally, $\mathbf{Y}_\tau$ may be constructed on another mesh $\mathcal{T}_\tau$ and with help of different trial functions.** Then

$$\|\nabla(\mathbf{u} - \mathbf{u_h})\| \leq \inf_{\mathbf{y_h} \in \mathbf{Y_h}} \{\|\nabla \mathbf{u_h} - \mathbf{y_h}\| + \mathbf{C}_\Omega \, \|\mathbf{div} \mathbf{y_h} + \mathbf{f}\|\}$$

The wider $\mathbf{Y_h} \subset \mathbf{H}(\Omega, \mathbf{div})$ the sharper is the upper bound.

Quadratic type functional

From the technical point of view it is better to square both parts of the estimate and apply minimization to a quadratic functional, namely

$$\|\nabla(\mathbf{u}-\mathbf{u_h})\|^2 \le \inf_{\mathbf{y_h}\in\mathbf{Y_h}} \Big\{(1+\beta)\|\nabla\mathbf{u_h}-\mathbf{y_h}\| + \\ + \mathbf{C_\Omega}\left(1+\frac{1}{\beta}\right)\|\mathbf{div}\mathbf{y_h}+\mathbf{f}\|^2\Big\}$$

Here, the positive parameter β can be also used to minimize the right–hand side.

Before going to more complicated problems where Deviation Majorants are derived by a more sophisticated theory, we observe several simple examples that nevertheless reflect key points of the above method.

Simple 1-D problem

$$
\begin{aligned}
(\,\alpha(x)\,u'\,)' &= f(x), \\
u(a) = 0, \quad & u(b) = u_b.
\end{aligned}
$$

It is equivalent to the variational problem

$$
J(v) = \int_a^b \left(\frac{1}{2}\alpha(x) \mid v' \mid^2 + f(x)v \right) dx.
$$

Assume that the coefficient α belongs to $\in L^\infty$ and bounded from below by a positive constant. Now

$$
V_0 + u_0 = \{ v \in H^1(a,b) \mid v(a) = 0,\, v(b) = u_b \}.
$$

Deviation Majorant

$$\mathcal{M}_{\oplus}(\mathbf{v},\beta,\mathbf{y}) \;=\; (1+\beta)\left(\int_{\mathbf{a}}^{\mathbf{b}} |\,\alpha\mathbf{v}'-\mathbf{y}\,|^2\,\mathbf{dx}+\frac{\mathbf{C}^2_{(\mathbf{a},\mathbf{b})}}{\beta}\int_{\mathbf{a}}^{\mathbf{b}}|\mathbf{y}'-\mathbf{f}|^2\right)\mathbf{dx}. \quad (65)$$

In this simple model, $\mathbf{u}$ can be presented in the form

$$\mathbf{u(x)} \;=\; \int_{\mathbf{a}}^{\mathbf{x}} \frac{\mathbf{1}}{\alpha(\mathbf{t})}\int_{\mathbf{a}}^{\mathbf{t}} \mathbf{f(z)dzdt} + \frac{\mathbf{x}}{\mathbf{b}}\left(\mathbf{u_b} - \int_{\mathbf{a}}^{\mathbf{b}} \frac{\mathbf{1}}{\alpha(\mathbf{t})}\int_{\mathbf{a}}^{\mathbf{t}} \mathbf{f(z)dzdt}\right).$$

what gives an opportunity to verify how error estimation methods work.

Approximations

Let $\mathbf{V_h}$ be made of piecewise–$\mathbf{P^1}$ continuous functions on uniform splittings of the interval and consider approximations of the following types:

- Galerkin approximations;
- Approximations very close to Galerkin (sharp);
- Approximations which are "good" but not Galerkin;
- Coarse (rough) approximations.

Our aim is to show that the Deviation Majorant can be effectively used as an error estimation instrument in all the above cases.

Computation of the Majorant

To find a sharp upper bound, we minimize $\mathcal{M}_\oplus$ with respect to $\mathbf{y}$ and β starting from the function $\mathbf{y_0} = \mathbf{G}(\mathbf{v}')$, where $\mathbf{G}$ is a simple averaging operator, e.g, defined by the relations

$$\mathbf{G}(\mathbf{v}')(\mathbf{x_i}) = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{v}'(\mathbf{x_i} - \mathbf{0}) + \mathbf{v}'(\mathbf{x_i} + \mathbf{0})),$$

By the quantity

$$\inf_{\beta > 0} \mathcal{M}_\oplus(\mathbf{v}, \beta, \mathbf{y_0}),$$

we obtain a coarse upper bound of the error. It is further improved by minimizing $\mathcal{M}_\oplus$ with respect to $\mathbf{y}$.

Example

Let $\alpha(x) = 1$, $\mathbf{f(x) = c}$, $\mathbf{a = 0}$, $\mathbf{b = 1}$, $\mathbf{u_b = 1}$, e.g., we consider the problem

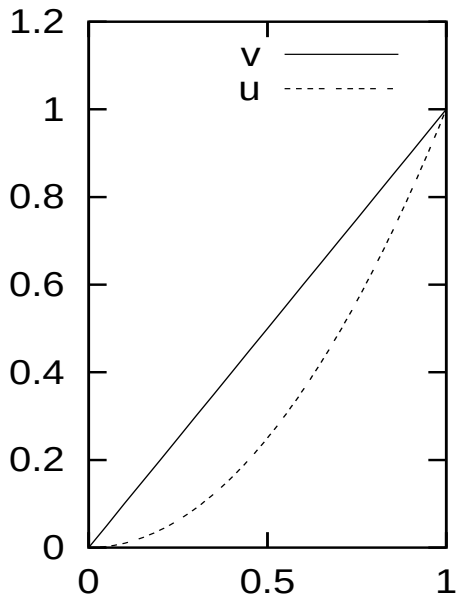
$$\mathbf{u'' = 2}, \quad \mathbf{u(0) = 0}, \ \mathbf{u(1) = 1}.$$

In this case, $\mathbf{C_{(a,b)} = 1/\pi}$ and

$$\mathbf{u = \frac{c}{2} x^2 + (1 - \frac{c}{2})x}, \quad \mathbf{u' = cx + 1 - \frac{c}{2}}.$$

Take a rough approximation $\mathbf{v = x}$. Then

$$\mathbf{\|(u - v)'\|^2 = \int_0^1 c^2(x - 0.5)^2 dx = c^2/12 \approx 0.083c^2}.$$



Exact solution and an approximation.

Various y give different upper bounds

(a) Take $\mathbf{y} = \mathbf{v}' = \mathbf{1}$, then the first term in

$$\mathcal{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{y}) \quad = \quad (\mathbf{1} \ + \ \beta) \left(\int_0^1 | \ \mathbf{v}' - \mathbf{y} \ |^2 \ \mathbf{dx} + \frac{\mathbf{1}}{\pi^2 \beta} \int_0^1 |\mathbf{y}' - \mathbf{f}|^2 \right) \mathbf{dx}.$$

vanishes and we have $\mathcal{M}_{\oplus} \longrightarrow \mathbf{c}^2/\pi^2 \approx \mathbf{0.101c}^2$; as $\beta \longrightarrow +\infty$. We see that this upper bound overestimates true error. Note that in this case, all sensible averagings of $v' = 1$ give exactly the same function: $\mathbf{G(1) = 1}$! Therefore,

$$\mathbf{G}(\mathbf{v}') \ - \ \mathbf{v}' \ \equiv \ \mathbf{0}$$

and formally ZZ indicator "does not see the error".

For the choice $\mathbf{y} = \mathbf{v}'$ the Majorant give a certain upper bound of the error (which is not so bad), but the integrand cannot indicate the distribution of local errors. Indeed, we have

$$\mathcal{M}_{\oplus} = \frac{1}{\pi^2} \int_0^1 \mathbf{c}^2 \mathbf{dx}.$$

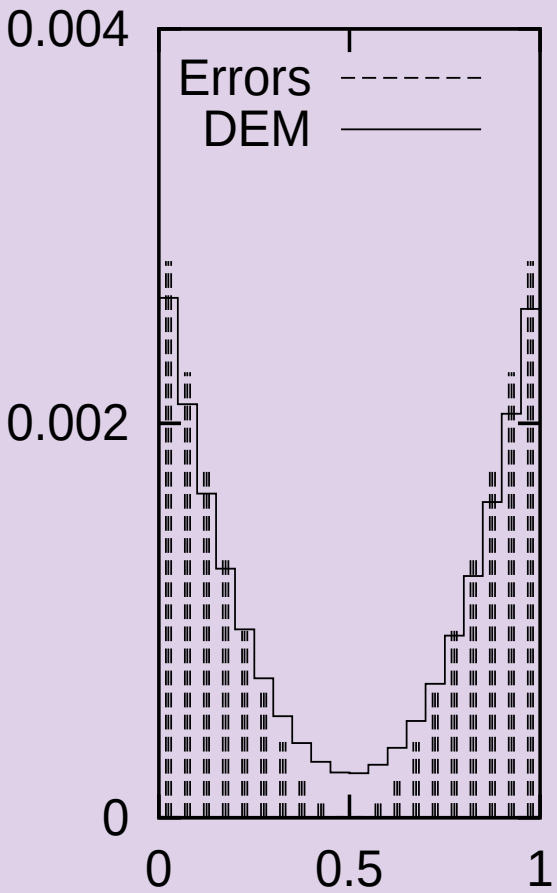
However, the integrand of the Majorant is a constant function, but the error is distributed in accordance with a parabolic law:

$$(\mathbf{u} - \mathbf{v})' = \mathbf{c}(\mathbf{x} - \mathbf{0.5})^2.$$

(b). Take $\mathbf{y = cx + 1 - c/2}$. Then, $\mathbf{y' = c}$ and the second term of the majorant vanishes. We have (for $\beta = 0$)

$$\mathcal{M}_{\oplus} = \int_0^1 \mathbf{c}^2(\mathbf{x} - \mathbf{1/2})^2 \mathbf{dx} = \mathbf{c}^2/\mathbf{12}.$$

We observe that both the global error and the error distribution are exactly reproduced. In real life computations such an "ideal" function $\mathbf{y}$ may be unattainable. However, the minimization makes the Majorant close to the sharp value. In this elementary example, we have minimized the Majorant on using piecewise affine approximations of $\mathbf{y}$ on 20 subintervals. The elementwise error distribution obtained as the result of this procedure is exposed on the next picture.



True errors and those computed by the Majorant.

To give further illustrations, we consider the functions

$$\mathbf{u}_\delta = \mathbf{u} + \delta\boldsymbol{\varphi},$$

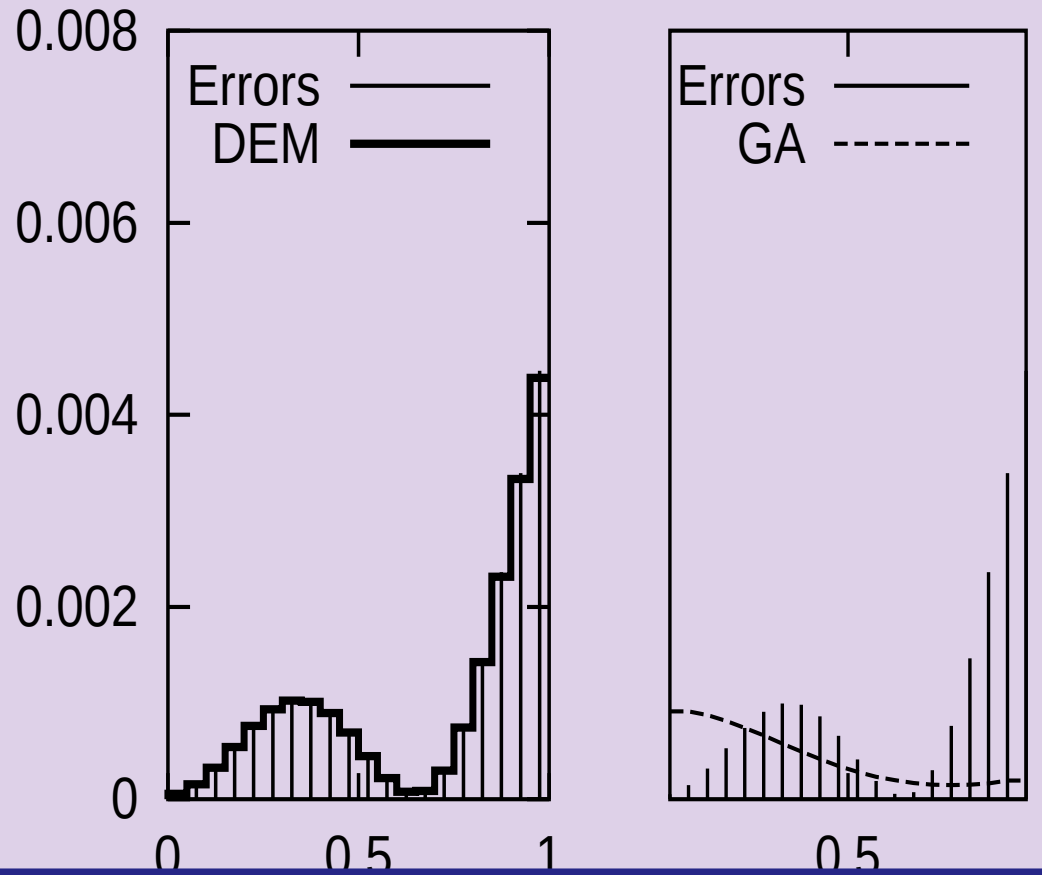
where δ is a number and $\boldsymbol{\varphi}$ is a certain function (perturbation).
Approximate solutions (whose errors are measured) are piecewise affine continuous interpolants of $\mathbf{u}_\delta$ defined on a uniform mesh with 20 subintervals. We take $\boldsymbol{\varphi} = \mathbf{x}\sin(\pi\mathbf{x})$ and $\delta = 0.1,\ 0.01,\ 0.001$, and 0.

Table:

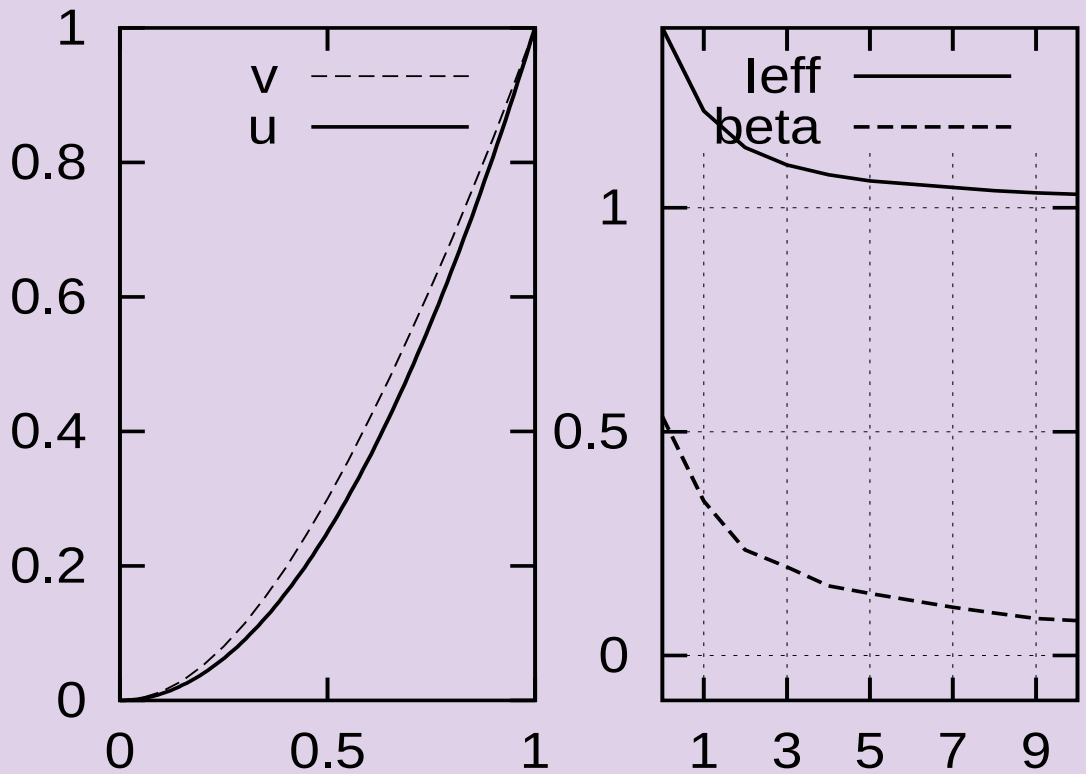
δ	e	$2\mathcal{M}_\oplus$	$2\mathcal{M}_\ominus$	$\mathrm{i}_{\mathrm{eff}}$	$\mathrm{i}_{\mathrm{esh}}$
0.1	0.019692	0.019743	0.019683	1.003	1.018
0.01	0.001022	0.001025	0.001013	1.003	1.011
0.001	0.000835	0.000839	0.000827	1.005	1.002
0	0.000833	0.000836	0.000825	1.004	1.002

In this experiment the Majorant was computed for $\frac{1}{2}\|\mathbf{e}\|^2$.

Error estimation for $\delta = 0.1$

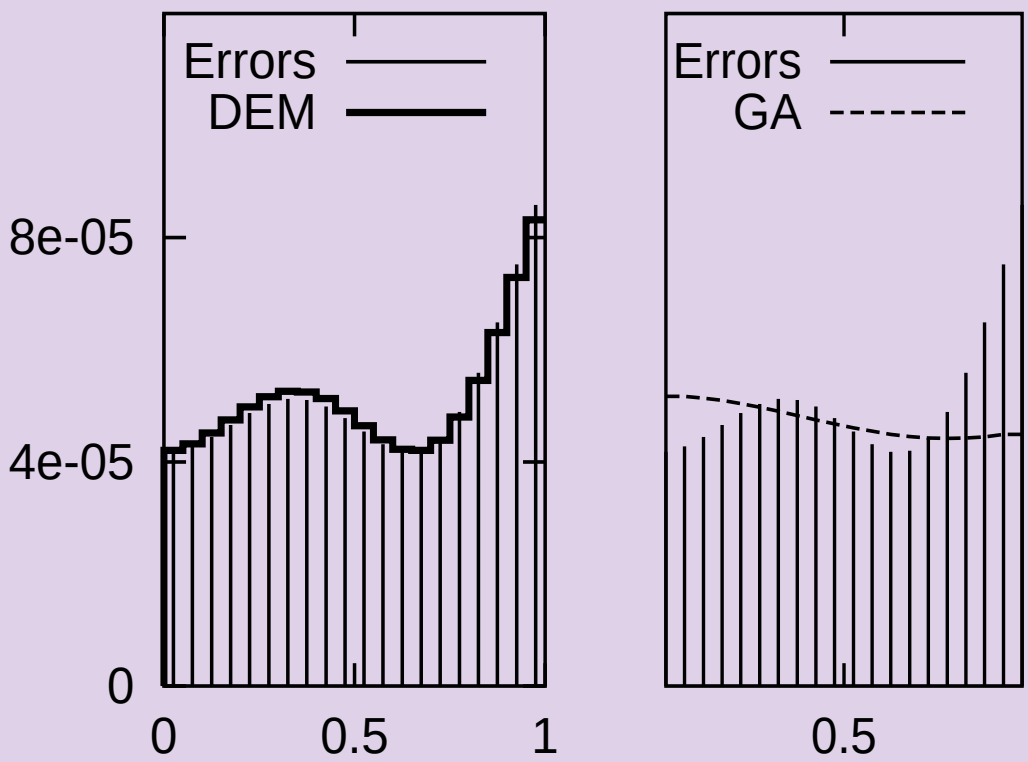


Functions **u**, **v** and $\mathbf{i_{eff}}$ for $\delta = 0.1$

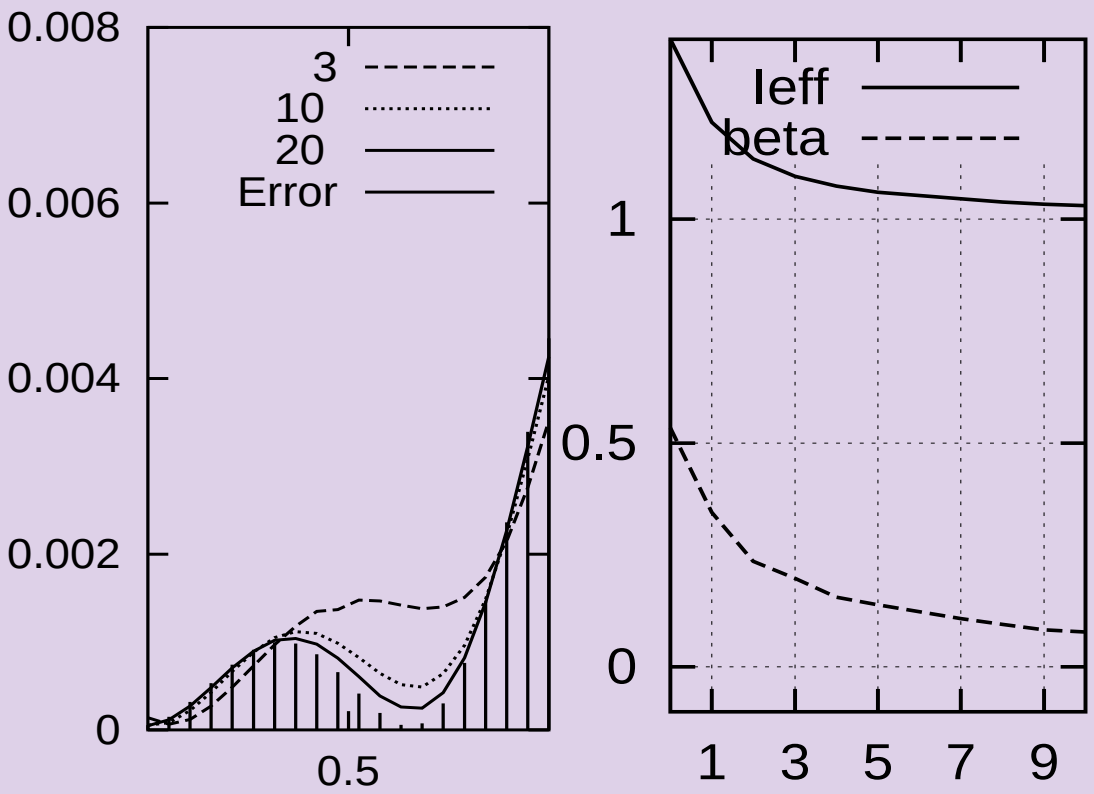


Error estimation for $\delta = 0.01$

A more precise approximation.

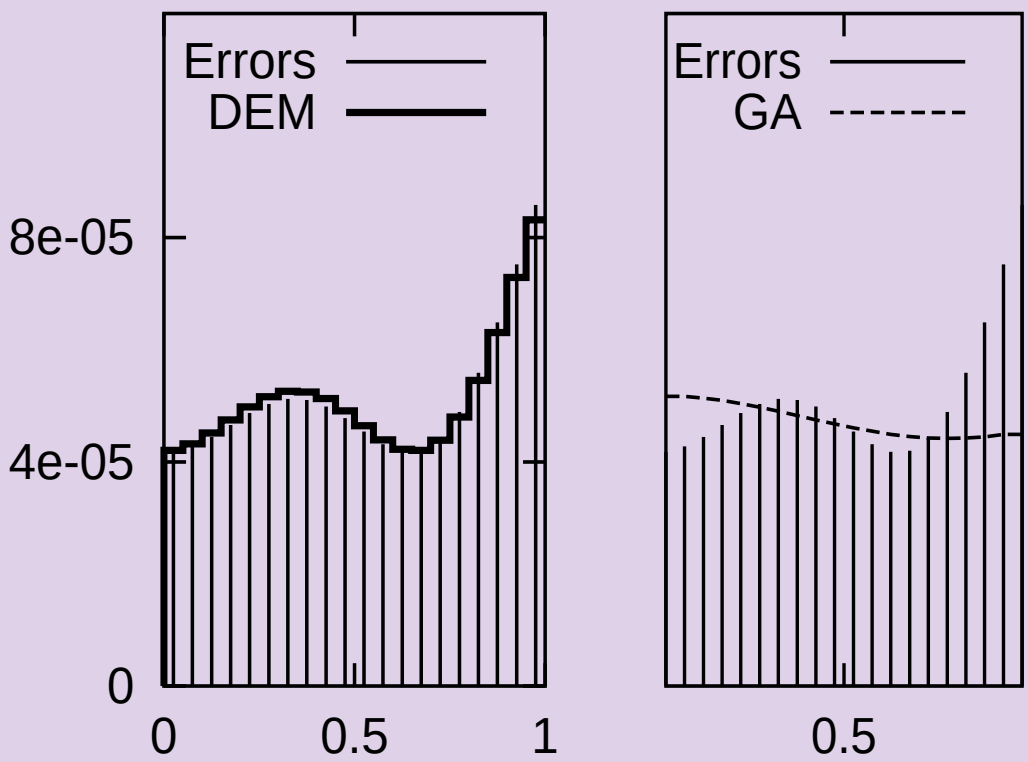


Functions $\mathbf{e}(\mathbf{y})$, β and $\mathbf{i_{eff}}$ for $\delta = 0.1$

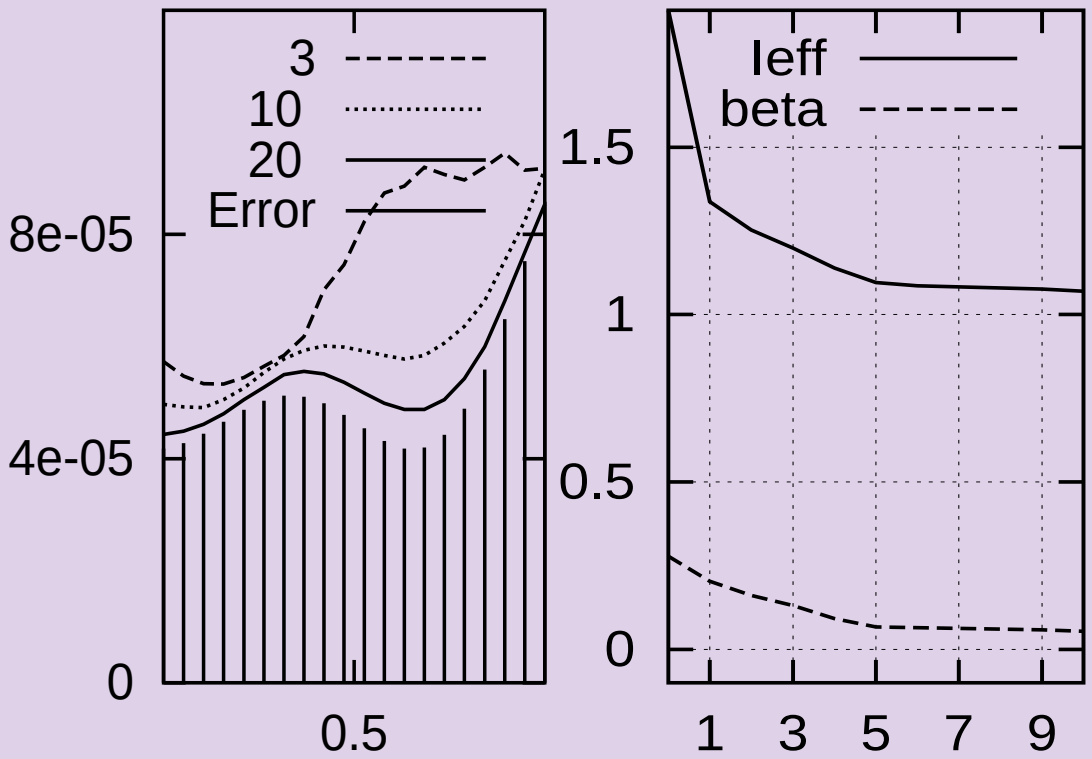


Error estimation for $\delta = 0.01$

A more precise approximation.

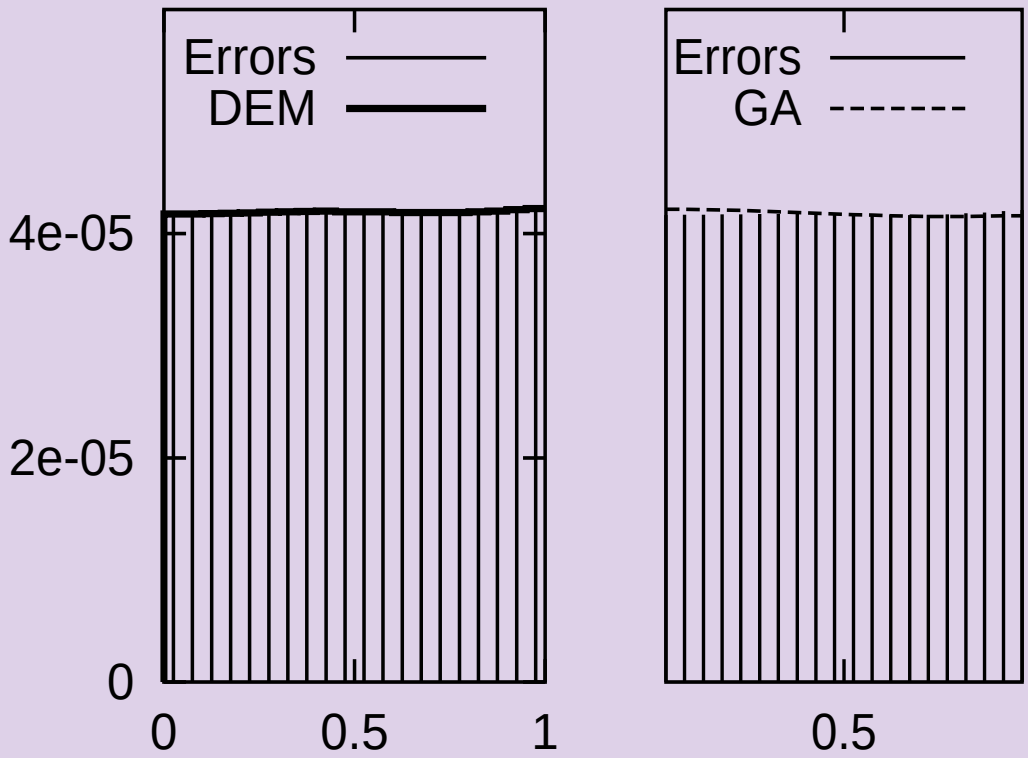


Functions $\mathbf{e}(\mathbf{y})$, β and $\mathbf{i_{eff}}$ for $\delta = 0.01$

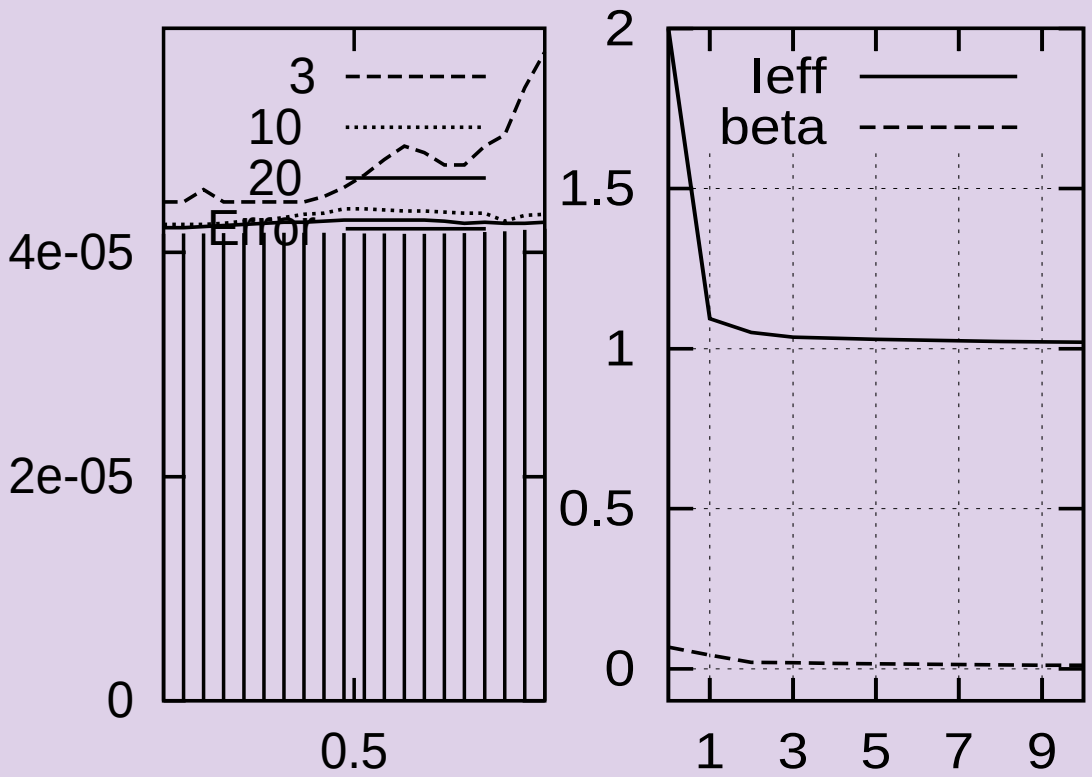


Error estimation for $\delta = 0.001$

Sharp approximation.

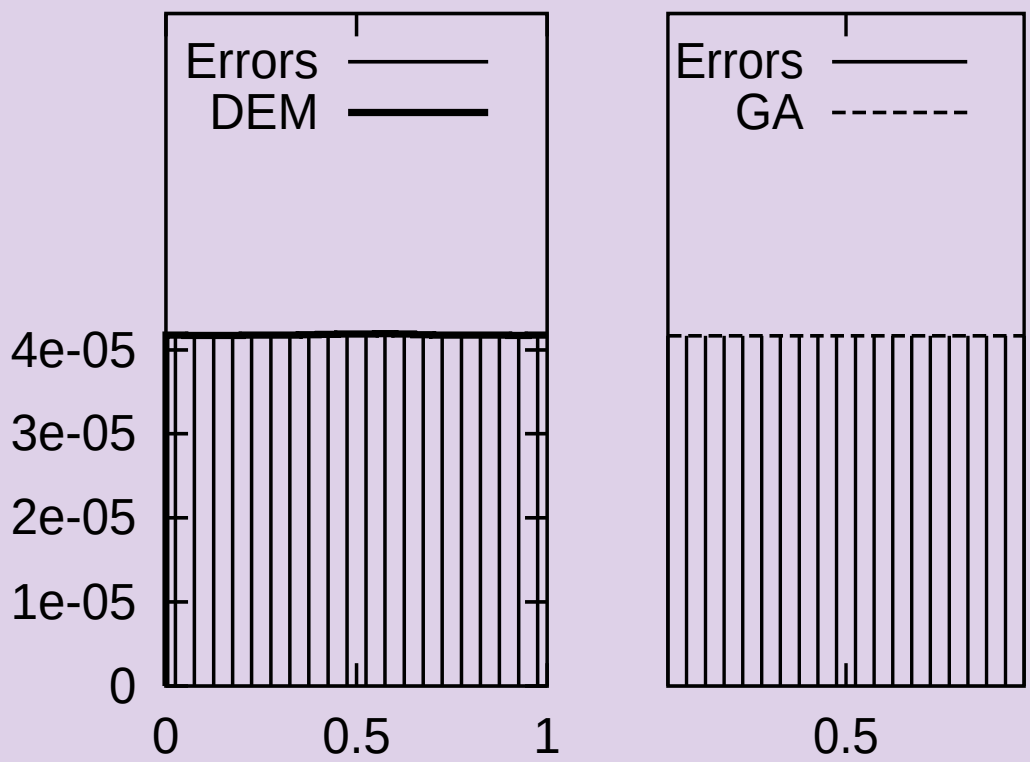


Functions $\mathbf{e}(\mathbf{y})$, β and $\mathbf{i_{eff}}$ for $\delta = 0.001$

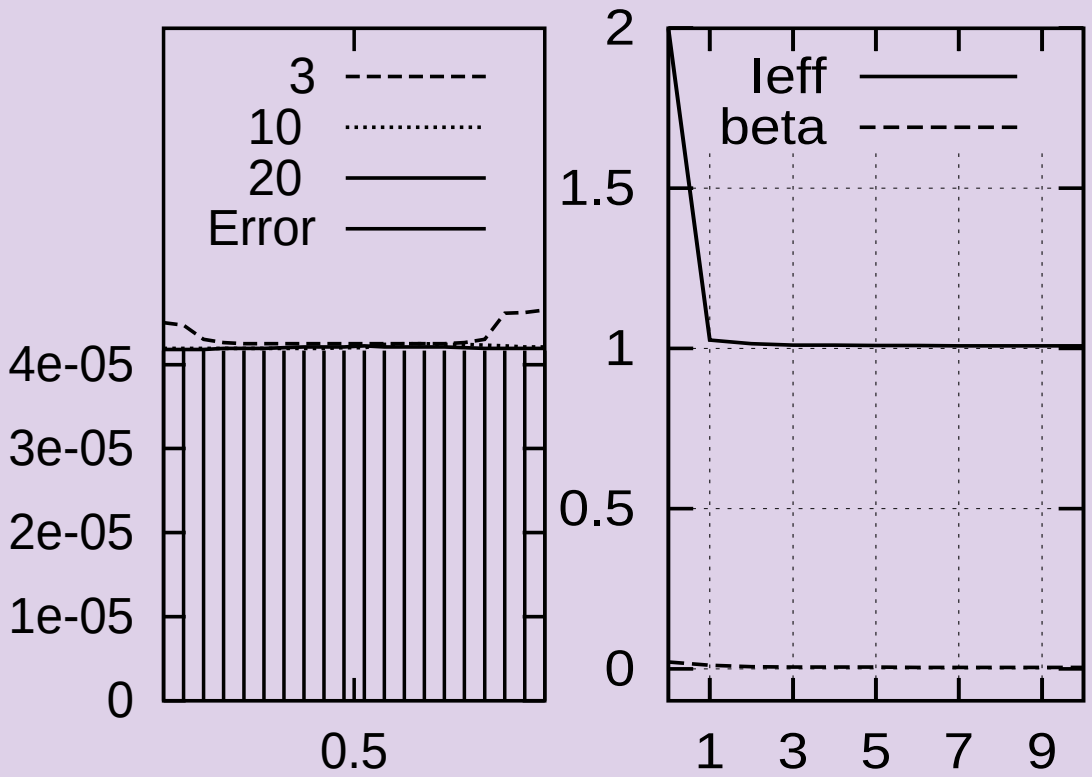


Error estimation for $\delta = 0$

Interpolant of the exact solution.



Functions $\mathbf{e}(\mathbf{y})$, β and $\mathbf{i_{eff}}$ for $\delta = 0$



Task 6

Apply the above theory to the problem

$$(\alpha \mathbf{u}')' = \mathbf{f},$$
$$\mathbf{u}(0) = \mathbf{0}, \quad \mathbf{u}(1) = \mathbf{b}$$

with your own α, $\mathbf{f}$, and $\mathbf{b}$. Compute approximate solutions and verify their accuracy along the same lines as in the example above.

Other examples

For problems with lower terms it is easy to obtain estimates without C_Ω.

$$\mathbf{\Delta u} - \varrho \mathbf{u} + \mathbf{f} = \mathbf{0}, \qquad \varrho > \mathbf{0},$$
$$\mathbf{u} = \mathbf{u_0} \quad \text{on } \partial\mathbf{\Omega}.$$

Such estimates can be derived by both *variational* and *non-variational* method. Let $\mathbf{w} \in \mathbf{V_0} := \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\mathbf{\Omega})$. We have

$$\int_{\mathbf{\Omega}} \nabla(\mathbf{u}-\mathbf{v}) \cdot \nabla \mathbf{w}\, \mathbf{dx} + \varrho \int_{\mathbf{\Omega}} (\mathbf{u}-\mathbf{v})\mathbf{w}\mathbf{dx} =$$
$$= \int_{\mathbf{\Omega}} (\mathbf{fw} - \nabla \mathbf{v} \cdot \nabla \mathbf{w})\mathbf{dx} - \varrho \int_{\mathbf{\Omega}} \mathbf{vw}\, \mathbf{dx}.$$

Use the integral identity for $\mathbf{y} \in \mathbf{H}(\mathbf{\Omega}, \mathbf{div})$:

$$\int_{\mathbf{\Omega}} (\nabla \mathbf{w} \cdot \mathbf{y} + \mathbf{w} \mathbf{div} \mathbf{y})\, \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

$$\int_{\mathbf{\Omega}} \nabla(\mathbf{u} - \mathbf{v}) \cdot \nabla \mathbf{w}\, \mathbf{dx} + \varrho \int_{\mathbf{\Omega}} (\mathbf{u} - \mathbf{v}) \mathbf{w} \mathbf{dx} =$$

$$\int_{\mathbf{\Omega}} (\mathbf{f} + \mathbf{div} \mathbf{y} - \varrho \mathbf{v}) \mathbf{w}\, \mathbf{dx} + \int_{\mathbf{\Omega}} (\mathbf{y} - \nabla \mathbf{v}) \cdot \nabla \mathbf{w})\, \mathbf{dx} \leq$$

$$\leq \|\mathbf{f} + \mathbf{div} \mathbf{y} - \varrho \mathbf{v}\| \|\mathbf{w}\| + \|\nabla \mathbf{v} - \mathbf{y}\| \|\nabla \mathbf{w}\|.$$

Set $\mathbf{w} = \mathbf{u} - \mathbf{v}$ and note that

$$\|\mathbf{f} + \mathbf{div}\mathbf{y} - \varrho\mathbf{v}\|\|\mathbf{u} - \mathbf{v}\| + \|\nabla\mathbf{v} - \mathbf{y}\|\|\nabla(\mathbf{u} - \mathbf{v})\| =$$
$$= \frac{\mathbf{1}}{\varrho}\|\mathbf{f} + \mathbf{div}\mathbf{y} - \varrho\mathbf{v}\|\varrho\|\mathbf{u} - \mathbf{v}\| + \|\nabla\mathbf{v} - \mathbf{y}\|\|\nabla(\mathbf{u} - \mathbf{v})\| \leq$$
$$\leq \left(\frac{\mathbf{1}}{\varrho^{\mathbf{2}}}\|\mathbf{f} + \mathbf{div}\mathbf{y} - \varrho\mathbf{v}\|^{\mathbf{2}} + \|\nabla\mathbf{v} - \mathbf{y}\|^{\mathbf{2}}\right)^{\mathbf{1/2}} \|\!|\!| \mathbf{u} - \mathbf{v} |\!|\!|$$

where

$$\|\!|\!| \mathbf{u} - \mathbf{v} |\!|\!|^{\mathbf{2}} = \int_{\mathbf{\Omega}} (|\nabla(\mathbf{u} - \mathbf{v})|^{\mathbf{2}} + \varrho|\mathbf{u} - \mathbf{v}|^{\mathbf{2}})\mathbf{dx}.$$

Then, we obtain the estimate

$$\| \mathbf{u} - \mathbf{v} \|^2 \leq \frac{\mathbf{1}}{\varrho^2} \|\mathbf{f} + \mathbf{div}\mathbf{y} - \varrho\mathbf{v}\|^2 + \|\nabla\mathbf{v} - \mathbf{y}\|^2$$

By the variational method this estimate was derived in 97'. Also, it readily follows from the general a posteriori framework
(see, e.g., S.R. Math. Comp. 2000).

This estimate has no $\mathbf{C_\Omega}$. However, in practice, it may give *big overestimation* if ϱ is small due to large penalty at the first term.

How Functional A Posteriori Estimate looks like for the problem

$$\mathbf{div A}\nabla \mathbf{u} + \mathbf{f} = \mathbf{0}, \quad \mathbf{u} = \mathbf{u_0} \quad \text{on } \partial\Omega$$

and for problems with boundary conditions of other types?

For the generalized diffusion equation it is as follows:

$$\| \nabla(\mathbf{u}-\mathbf{v}) \|^2 \leq \sqrt{\mathbf{D}(\nabla\mathbf{v},\mathbf{y})} + \mathbf{C}(\Omega,\mathbf{A})\|\mathbf{div}\mathbf{y}+\mathbf{f}\|,$$

where

$$\begin{aligned}
\| \eta \|^2 := & \int_\Omega \mathbf{A}\nabla\eta\cdot\nabla\eta\,\mathbf{dx}, \\
\mathbf{D}(\nabla\mathbf{v},\mathbf{y}) := & \int_\Omega (\mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{v}+\mathbf{A}^{-1}\mathbf{y}\cdot\mathbf{y}-2\nabla\mathbf{v}\cdot\mathbf{y})\mathbf{dx} = \\
& = \int_\Omega (\mathbf{A}\nabla\mathbf{v}-\mathbf{y})\cdot(\nabla\mathbf{v}-\mathbf{A}^{-1}\mathbf{y})\mathbf{dx}, \\
\|\mathbf{w}\| \leq & \quad \mathbf{C}(\Omega,\mathbf{A}) \| \nabla\mathbf{w} \| .
\end{aligned}$$

How such estimates are derived we will discuss in the next Lecture.

Lectute 4

We expose the general approach to deriving two-sided functional estimates of the deviations from exact solutions of linear elliptic type problems having the operator form

$$\mathbf{\Lambda}^* \mathcal{A} \mathbf{\Lambda} \mathbf{u} + \ell = \mathbf{0}$$

where $\mathbf{\Lambda}$ and $\mathcal{A}$ are linear bounded operators and $\mathcal{A}$ is symmetric and positive definite.

Lecture plan

- Two–sided a posteriori estimates for linear elliptic type problems;
- Properties: computability, consistency, reliability;
- Relationships with other error estimation methods;
- Diffusion equation with Dirichlét boundary conditions;
- Diffusion equation with Neumann boundary conditions;
- Diffusion equation with mixed boundary conditions;
- Linear elasticity with mixed boundary conditions;

Problem in the abstract form

Many problems can be presented in the following form: **find $\mathbf{u} \in \mathbf{V_0} + \mathbf{u_0}$ such that**

$$(\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda w}) + \langle \ell, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0}. \tag{66}$$

Here $\mathbf{V_0}$ is a subspace of a reflexive Banach space $\mathbf{V}$,
e.g., $\mathbf{V} = \mathbf{H^1}$, $\mathbf{V_0} = \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}$.
$\mathbf{\Lambda} : \mathbf{V} \longrightarrow \mathbf{U}$ is a bounded linear operator, e.g. $\mathbf{\Lambda} = \nabla$.
$\mathbf{U}$ is a Hilbert space with scalar product $(\cdot, \cdot)$ and norm $\| \cdot \|$,
e.g., $\mathbf{U} = \mathbf{L^2}$.
$\ell \in \mathbf{V_0^*}$, is a linear functional in the dual space, e.g., in $\mathbf{H^{-1}}$. In general, we may assume that

$$\langle \ell, \mathbf{w} \rangle = (\mathbf{f}, \mathbf{w}) + (\mathbf{g}, \mathbf{\Lambda w}).$$

$\mathcal{A} \in \mathcal{L}(\mathbf{U}, \mathbf{U})$ is a self-adjoint operator.

Assumptions

We assume that

$$\mathbf{V} \quad \text{is compactly embedded in} \quad \mathbf{U} \tag{67}$$

and the operators $\mathbf{\Lambda}$ and $\mathcal{A}$ satisfy the relations

$$\mathbf{c_1}\|\mathbf{y}\|^2 \leq (\mathcal{A}\mathbf{y}, \mathbf{y}) \leq \mathbf{c_2}\|\mathbf{y}\|^2, \qquad \forall \mathbf{y} \in \mathbf{U}, \tag{68}$$

$$\|\mathbf{\Lambda w}\| \geq \mathbf{c_3}\|\mathbf{w}\|_{\mathbf{V}}, \quad \forall \mathbf{w} \in \mathbf{V_0}, \tag{69}$$

For our analysis, it is convenient to introduce two more norms:

$$\| \mathbf{y} \| := (\mathcal{A}\mathbf{y}, \mathbf{y})^{1/2}, \qquad \| \mathbf{y} \|_* = (\mathcal{A}^{-1}\mathbf{y}, \mathbf{y})^{1/2},$$

where $\mathcal{A}^{-1} : \mathbf{U} \to \mathbf{U}$
is the operator inverse to $\mathcal{A}$. The respective spaces $\mathbf{Y}$ and $\mathbf{Y}^*$ contain elements of $\mathbf{U}$ equipped with the norms $\| \cdot \|$ and $\| \cdot \|_*$, respectively.
Let $\mathbf{\Lambda}^*$ be the operator conjugate to $\mathbf{\Lambda}$, i.e.,

$$(\mathbf{y}, \mathbf{\Lambda w}) = \langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle, \qquad \forall \mathbf{w} \in \mathbf{V_0}. \tag{70}$$

Getting a posteriori estimates by transformations of integral identities

For a detailed exposition see
S.Repin. Two-sided estimates of deviation from exact solutions of uniformly elliptic equations, Proc. St. Petersburg Math. Society, IX(2001), pp. 143–171, translation in Amer. Math. Soc. Transl. Ser. 2, 209, Amer. Math. Soc., Providence, RI, 2003.

Let $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$ be an approximation.

$$(\mathcal{A}\mathbf{\Lambda}(\mathbf{u} - \mathbf{v}), \mathbf{\Lambda w}) + \langle \ell, \mathbf{w} \rangle + (\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda w}) = \mathbf{0} \qquad \forall \mathbf{w} \in \mathbf{V_0}\,,$$

Set $\mathbf{w} = \mathbf{u} - \mathbf{v}$.

$$\| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2 = | \langle \ell, \mathbf{u} - \mathbf{v} \rangle + (\mathcal{A}\mathbf{\Lambda}\mathbf{v}, \mathbf{\Lambda}(\mathbf{u} - \mathbf{v})) | .$$

By

$$(\mathbf{y}, \mathbf{\Lambda}\mathbf{w}) = \langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle, \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

we have

$$\| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2 = | \langle \ell, \mathbf{u} - \mathbf{v} \rangle + (\mathcal{A}\mathbf{\Lambda}\mathbf{v} - \mathbf{y}, \mathbf{\Lambda}(\mathbf{u} - \mathbf{v})) + \langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{u} - \mathbf{v} \rangle) | .$$

Therefore, we find that

$$\| \mathbf{\Lambda}(\mathbf{u}-\mathbf{v}) \|^2 \leq | \langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{u}-\mathbf{v} \rangle | + | (\mathcal{A}\mathbf{\Lambda}\mathbf{v} - \mathbf{y}, \mathbf{\Lambda}(\mathbf{u}-\mathbf{v})) | \leq$$
$$\leq \| \ell + \mathbf{\Lambda}^* \mathbf{y} \| \, \| \mathbf{\Lambda}(\mathbf{u}-\mathbf{v}) \| + \| \mathcal{A}\mathbf{\Lambda}\mathbf{v} - \mathbf{y} \|_* \, \| \mathbf{\Lambda}(\mathbf{u}-\mathbf{v}) \| ,$$

where

$$\| \mu \| := \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\langle \mu, \mathbf{w} \rangle}{\| \mathbf{\Lambda} \mathbf{w} \|}$$

denotes the norm of the functional $\mu : \mathbf{V_0} \to \mathbb{R}$.
To prove that the value of $\| \ell + \mathbf{\Lambda}^* \mathbf{y} \|$ is finite we note that

$$\langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle \leq \|\ell\|_{\mathbf{V_0}^*} \|\mathbf{w}\|_{\mathbf{V}} + \|\mathbf{y}\| \|\mathbf{\Lambda}\mathbf{w}\| \leq \left(\mathbf{c_3}^{-1} \|\ell\|_{\mathbf{V_0}^*} + \|\mathbf{y}\| \right) \|\mathbf{\Lambda}\mathbf{w}\| \leq$$
$$\leq \mathbf{c_1}^{-1/2} \left(\mathbf{c_3}^{-1} \|\ell\|_{\mathbf{V_0}^*} + \|\mathbf{y}\| \right) \| \mathbf{\Lambda}\mathbf{w} \| .$$

General estimate

As a result we obtain the general form of a functional a posteriori estimate for an elliptic type problem:

$$\| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \| \leq \mathbf{[} \ell + \mathbf{\Lambda}^* \mathbf{y} \mathbf{]} + \| \mathcal{A}\mathbf{\Lambda}\mathbf{v} - \mathbf{y} \|_* \ . \tag{71}$$

Denote

$$\mathbf{D}(\mathbf{y}_1, \mathbf{y}_2) := \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}\mathbf{y}_1, \mathbf{y}_1) + \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}^{-1}\mathbf{y}_2, \mathbf{y}_2) - (\mathbf{y}_1, \mathbf{y}_2) =$$

$$= \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}(\mathbf{y}_1 - \mathcal{A}^{-1}\mathbf{y}_2), \mathbf{y}_1 - \mathcal{A}^{-1}\mathbf{y}_2) = \frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{y}_1 - \mathcal{A}^{-1}\mathbf{y}_2 \|^2 =$$

$$= \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}^{-1}(\mathbf{y}_2 - \mathcal{A}\mathbf{y}_1), \mathbf{y}_2 - \mathcal{A}\mathbf{y}_1) = \frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{y}_2 - \mathcal{A}\mathbf{y}_1 \|_*^2$$

Then

$$\| \mathcal{A}\mathbf{\Lambda v} - \mathbf{y} \|_*^2 = \mathbf{2D}(\mathbf{\Lambda v}, \mathbf{y})$$

and we obtain

$$\| \Lambda(u - v) \| \leq \; \| \ell + \Lambda^* y \| \; + \; \sqrt{2D(\Lambda v, y)}\,.$$

Square both sides and use Young's inequality

$$\| \Lambda(u-v) \|^2 \leq \left(1+\tfrac{1}{\beta}\right) \| \ell+\Lambda^* y \|^2 + 2(1+\beta) D(\Lambda v, y). \qquad (72)$$

This estimate holds for any $y \in Y^*$ and $\beta > 0$. Denote its right–hand side by $2\mathcal{M}_{\oplus}(v, \beta, y)$

$2\mathcal{M}_{\oplus}(\mathbf{v},\beta,\mathbf{y})$ is a sharp upper bound

Proposition

For any $\beta > 0$ there exists $\mathbf{y} \in \mathbf{Y}^*$ such that

$$2\mathcal{M}_{\oplus}(\beta,\mathbf{v},\mathbf{y}) = \| \mathbf{\Lambda}(\mathbf{u}-\mathbf{v}) \|^2 \ .$$

Proof. Set $\mathbf{y_1} = \frac{1}{1+\beta}(\mathbf{p} + \beta\mathcal{A}\mathbf{\Lambda v})$ where $\mathbf{p} = \mathcal{A}\mathbf{\Lambda u}$. Note that $\langle \ell + \mathbf{\Lambda}^* \mathbf{y_1}, \mathbf{w} \rangle = (-\mathbf{p} + \mathbf{y_1}, \mathbf{\Lambda w})$.

$$\| \ell + \mathbf{\Lambda}^* \mathbf{y_1} \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{(\mathbf{y_1} - \mathbf{p}, \mathbf{\Lambda w})}{\| \mathbf{\Lambda w} \|} =$$

$$= \frac{\beta}{1+\beta} \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{(\mathcal{A}\mathbf{\Lambda}(\mathbf{v}-\mathbf{u}), \mathbf{\Lambda w})}{\| \mathbf{\Lambda w} \|} = \frac{\beta}{1+\beta} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \| ,$$

Similarly

$$\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}_1) = \frac{1}{2(1+\beta)^2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 .$$

Therefore,

$$\left(1+\frac{1}{\beta}\right) \| \ell + \mathbf{\Lambda}^* \mathbf{y} \|^2 + 2(1+\beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) =$$

$$\left(1+\frac{1}{\beta}\right) \left(\frac{\beta}{1+\beta} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|\right)^2 + 2(1+\beta)\left(\frac{1}{2(1+\beta)^2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2\right) =$$

$$= \frac{\beta}{1+\beta} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 + \frac{1}{1+\beta} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 = \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 .$$

We replace $\mathbf{[}\, \ell + \Lambda^* \mathbf{y} \,\mathbf{]}$ by the norm in a Hilbert space $\mathbf{U}$ provided that ℓ belongs to a narrower set. Assume that

$$\ell \in \mathbf{U} \subset \mathbf{V}_0^*,$$
$$\mathbf{y} \in \mathbf{Q}^* := \{\mathbf{z}^* \in \mathbf{Y}^* | \qquad \Lambda^* \mathbf{z}^* \in \mathbf{U}\}.$$

Note that $\mathbf{Q}^*$ can be endowed with the norm

$$\|\mathbf{y}\|^2_{\mathbf{Q}^*} := \|\mathbf{y}\|^2_* + \|\Lambda^* \mathbf{y}\|^2_{\mathbf{U}}.$$

If $\ell \in \mathbf{U}$, then $\mathbf{Q}^*$ contains the exact solution p of the dual problem! This fact is important for the proof of the sharpness of the Majorant.

Then

$$\langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = (\ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w}) \quad \mathbf{w} \in \mathbf{V_0}.$$

$$\mathbf{|\!|\!|}\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,\mathbf{|\!|\!|} = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle}{\| \mathbf{\Lambda w} \|} \leq \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\| \ell + \mathbf{\Lambda}^* \mathbf{y} \| \, \| \mathbf{w} \|}{\| \mathbf{\Lambda w} \|} \leq$$

$$\leq \| \ell + \mathbf{\Lambda}^* \mathbf{y} \| \mathbf{c}_1^{-1} \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\| \mathbf{w} \|}{\| \mathbf{\Lambda w} \|} \leq \mathbf{c}_1^{-1} \mathbf{c}_3^{-1} \| \ell + \mathbf{\Lambda}^* \mathbf{y} \|.$$

Denote $\mathbf{c}^2 = \mathbf{c}_1^{-2} \mathbf{c}_3^{-2}$.

Computable Majorant of the deviation

Now, the Majorant $\mathcal{M}_\oplus$ is replaced by $\mathbf{M}_\oplus$, namely we arrive at the estimate

$$\frac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 \leq \mathbf{M}_\oplus(\mathbf{v},\beta,\mathbf{y}) := \\ := (1+\beta)\mathbf{D}(\mathbf{\Lambda v},\mathbf{y}) + \frac{1+\beta}{2\beta}\mathbf{c}^2\|\ell + \mathbf{\Lambda}^*\mathbf{y}\|^2. \quad (73)$$

Variational Method

Problem $\mathcal{P}$. Find $\mathbf{u} \in \mathbf{V_0} + \mathbf{u_0}$ such that

$$\mathbf{J}(\mathbf{u}) = \inf_{\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}} \mathbf{J}(\mathbf{u}) := \inf \mathcal{P},$$

where

$$\mathbf{J}(\mathbf{v}) = \frac{\mathbf{1}}{\mathbf{2}} \parallel\!\!| \mathbf{\Lambda v} |\!\!\parallel^2 + \langle \ell, \mathbf{v} \rangle.$$

Lagrangian

On the set $(\mathbf{V_0} + \mathbf{u_0}) \times \mathbf{Y}^*$, we define the Lagrangian

$$\mathbf{L}(\mathbf{v}, \mathbf{y}) = (\mathbf{y}, \mathbf{\Lambda v}) - \frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{y} \|^2 + \langle \ell, \mathbf{v} \rangle$$

and the functional

$$\mathbf{I}^*(\mathbf{y}) = \inf_{\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}} \mathbf{L}(\mathbf{v}, \mathbf{y}) = \begin{cases} (\mathbf{y}, \mathbf{\Lambda u_0}) - \frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{y} \|_*^2 + \langle \ell, \mathbf{u_0} \rangle, & \mathbf{y} \in \mathbf{Q}_\ell^*, \\ -\infty, & \mathbf{y} \notin \mathbf{Q}_\ell^*, \end{cases}$$

where $\mathbf{Q}_\ell^* := \{ \mathbf{y} \in \mathbf{Y}^* \mid (\mathbf{y}, \mathbf{\Lambda w}) + \langle \ell, \mathbf{w} \rangle = \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0} \}$.
Note that since

$$(\mathbf{y}, \mathbf{\Lambda}(\mathbf{u_0} + \mathbf{w}) + \langle \ell, (\mathbf{u_0} + \mathbf{w}) \rangle = (\mathbf{y}, \mathbf{\Lambda u_0}) + \langle \ell, \mathbf{u_0} \rangle$$

we see that $\mathbf{I}^*$ does not depend on the form of $\mathbf{u_0}$ inside Ω.

The problem dual to $\mathcal{P}$ is as follows.

Problem $\mathcal{P}^*$. Find $\mathbf{p} \in \mathbf{Q}_\ell^*$ such that

$$\mathsf{I}^*(\mathbf{p}) = \sup_{\mathbf{y} \in \mathbf{Q}_\ell^*} \mathsf{I}^*(\mathbf{y}) := \sup \mathcal{P}^* \leq \inf \mathcal{P}.$$

The minimizer $\mathbf{u}$ and the maximizer $\mathbf{p}$ satisfy the conditions

$$(\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda w}) + \langle \ell, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0},$$
$$(\mathbf{\Lambda u_0} - \mathcal{A}^{-1}\mathbf{p}, \mathbf{y}) = \mathbf{0}, \quad \forall \mathbf{y} \in \mathbf{Q_0^*},$$

where $\mathbf{Q_0^*} := \left\{ \mathbf{y} \in \mathbf{Y}^* \middle| (\mathbf{y}, \mathbf{\Lambda w}) = \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0} \right\}.$

We see that $\mathcal{A}\mathbf{\Lambda u} \in \mathbf{Q}_\ell^*$,.

Take

$$\mathsf{I}^*(\mathcal{A}\mathbf{\Lambda u}) = (\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda u_0}) - \frac{1}{2} \parallel \mathcal{A}\mathbf{\Lambda u}^2 \parallel + \langle \ell, \mathbf{u_0} \rangle.$$

Recall that the dual functional does not depend on $\mathbf{u_0}$ inside Ω. Therefore, we set $\mathbf{u_0} = \mathbf{u}$ and observe that

$$\mathsf{I}^*(\mathcal{A}\mathbf{\Lambda u}) = (\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda u}) - \frac{1}{2} \parallel \mathcal{A}\mathbf{\Lambda u} \parallel_*^2 + \langle \ell, \mathbf{u} \rangle \leq \sup \mathcal{P}^*.$$

Since $\parallel \mathcal{A}\mathbf{\Lambda u} \parallel_*^2 = (\mathcal{A}^{-1}\mathcal{A}\mathbf{\Lambda u}, \mathcal{A}\mathbf{\Lambda u}) = \parallel \mathbf{\Lambda u} \parallel^2$, we see that

$$\mathsf{I}^*(\mathcal{A}\mathbf{\Lambda u}) = \mathsf{J}(\mathbf{u}) = \inf \mathcal{P}$$

Thus

$$\sup \mathcal{P}^* = \inf \mathcal{P}$$

The relation $\mathsf{I}^*(\mathbf{p}) = \mathsf{J}(\mathbf{u})$ means that

$$(\mathbf{p}, \mathbf{\Lambda u}) - \frac{1}{2} \| \mathbf{p} \|_*^2 + \langle \ell, \mathbf{u} \rangle = \frac{1}{2} \| \mathbf{\Lambda u} \|^2 + \langle \ell, \mathbf{u} \rangle,$$

which is equivalent to the relation

$$\mathsf{D}(\mathbf{\Lambda u}, \mathbf{p}) = \frac{1}{2} \| \mathbf{\Lambda u} \|^2 + \frac{1}{2} \| \mathbf{p} \|_*^2 - (\mathbf{p}, \mathbf{\Lambda u}) = \mathbf{0}.$$

From the above we see that $\mathbf{\Lambda u}$ and $\mathbf{p}$ are joined by a certain relation:

$$\mathbf{p} = \mathcal{A} \mathbf{\Lambda u}$$

This is the so–called **duality relation** for the pair $(\mathbf{u}, \mathbf{p})$.

Let $v \in V_0 + u_0$ and $\mathbf{y} \in \mathbf{Y}^*$ be some approximations of $\mathbf{u}$ and $\mathbf{p}$, respectively. Our goal is to obtain two-sided estimates of the quantities $\| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|$ and $\| \mathbf{y} - \mathbf{p} \|_*$ that are norms of **deviations** from the exact solutions $\mathbf{u}$ and $\mathbf{p}$. First, we establish the following basic result.

Theorem

For any $v \in \mathbf{V_0} + \mathbf{u_0}$ *and* $\mathbf{q} \in \mathbf{Q}^*_\ell$,

$$\| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^2 + \| \mathbf{q} - \mathbf{p} \|^2_* = 2\,(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q})), \tag{74}$$

$$\| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^2 + \| \mathbf{q} - \mathbf{p} \|^2_* = 2\,\mathbf{D}(\mathbf{\Lambda v}, \mathbf{q}). \tag{75}$$

Proof

By the stationarity relations, we have

$$\begin{aligned}\tfrac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 &= \frac{1}{2}(\mathcal{A}\mathbf{\Lambda}(\mathbf{v}-\mathbf{u}), \mathbf{\Lambda}(\mathbf{v}-\mathbf{u})) = \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) + \\ &+(\mathcal{A}\mathbf{\Lambda}\mathbf{u}, \mathbf{\Lambda}(\mathbf{u}-\mathbf{v})) + \langle \ell, \mathbf{u}-\mathbf{v}\rangle = \\ &= \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}).\end{aligned}$$

Analogously

$$\begin{aligned}\frac{1}{2} \| \mathbf{q}-\mathbf{p} \|_*^2 &= \frac{1}{2}(\mathcal{A}^{-1}(\mathbf{p}-\mathbf{q}), \mathbf{p}-\mathbf{q}) = \\ &= \mathbf{I}^*(\mathbf{p}) - \mathbf{I}^*(\mathbf{q}) - (\mathbf{\Lambda}\mathbf{u_0} - \mathcal{A}^{-1}\mathbf{p}, \mathbf{p}-\mathbf{q}) = \mathbf{I}^*(\mathbf{p}) - \mathbf{I}^*(\mathbf{q}).\end{aligned}$$

Since $\mathbf{J}(\mathbf{u}) = \mathbf{I}^*(\mathbf{p})$, we sum two relations and obtain (74). For $\mathbf{q} \in \mathbf{Q}^*_\ell$ the difference $\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q})$ is equal to $\mathbf{D}(\mathbf{\Lambda}\mathbf{v}, \mathbf{q})$, so that (75) follows from (74).

The estimates (74) and (75) are valid only for $\mathbf{q} \in \mathbf{Q}^*_\ell$, which poses some difficulties. Below it is shown how we can overcome this drawback. First, we establish one subsidiary result.

Theorem

Let $\mathbf{q} \in \mathbf{Q}^*_\ell$, $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$, $\beta \in \mathbb{R}_+$, *and* $\mathbf{y} \in \mathbf{Y}^*$. *Then*

$$\mathbf{J(v)} - \mathbf{I}^*\mathbf{(q)} \quad \leq \quad \mathbf{(1}+\beta\mathbf{)D(\Lambda v, y)} + \frac{\mathbf{1}+\beta}{\mathbf{2}\beta} \quad \| \quad \mathbf{q} - \mathbf{y} \quad \|^{\mathbf{2}}_* \quad . \quad (76)$$

Note that

$$\mathbf{D(\Lambda v, y)} = \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda v}) + \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}^{-\mathbf{1}}\mathbf{p}, \mathbf{p}) - (\mathbf{y}, \mathbf{\Lambda u}) =$$
$$= (\mathcal{A}\mathbf{\Lambda v} - \mathbf{y}), \mathbf{\Lambda v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y}) =$$
$$= (\mathcal{A}(\mathbf{\Lambda v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y}, \mathbf{\Lambda v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y}) = \| \mathbf{\Lambda v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y} \|.$$

Proof

For any $\mathbf{y} \in \mathbf{Y}^*$, we have

$$\mathbf{J(v)} - \mathbf{I}^*\mathbf{(q)} = \frac{\mathbf{1}}{\mathbf{2}}\Big(\| \mathbf{\Lambda v} \|^{\mathbf{2}} + \| \mathbf{y} \|^{\mathbf{2}}_* \Big) + \\ + \frac{\mathbf{1}}{\mathbf{2}} \big(\| \mathbf{q} \|^{\mathbf{2}}_* - \| \mathbf{y} \|^{\mathbf{2}}_* \big) - \mathbf{(\Lambda u_0, q)} + \langle \ell, \mathbf{v - u_0} \rangle .$$

Since $\langle \ell, \mathbf{v - u_0} \rangle = \mathbf{(q, \Lambda(u_0 - v))}$, we find that

$$\mathbf{J(v)} - \mathbf{I}^*\mathbf{(q)} = \frac{\mathbf{1}}{\mathbf{2}} \left(\| \mathbf{\Lambda v} \|^{\mathbf{2}} + \| \mathbf{y} \|^{\mathbf{2}}_* \right) + \frac{\mathbf{1}}{\mathbf{2}} \left(\| \mathbf{q} \|^{\mathbf{2}}_* - \| \mathbf{y} \|^{\mathbf{2}}_* \right) - \mathbf{(q, \Lambda v)} = \\ = \mathbf{D(\Lambda v, y)} + \mathbf{\left(y - q, \Lambda v - \mathcal{A}^{-1} y\right)} + \frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{q - y} \|^{\mathbf{2}}_* .$$

This relation yields (76) if we use the Young's inequality

$$\mathbf{2} \left(\mathbf{y - q, \Lambda v - \mathcal{A}^{-1} y} \right) \le \beta \| \mathbf{\Lambda v - \mathcal{A}^{-1} y} \|^{\mathbf{2}} + \beta^{-1} \| \mathbf{y - q} \|^{\mathbf{2}}_* .$$

Another form of the estimate

Introduce the quantity

$$\mathbf{d}_\ell^2(\mathbf{y}) := \inf_{\mathbf{q}\in\mathbf{Q}_\ell^*} \| \mathbf{q}-\mathbf{y} \|_*^2,$$

which is the distance to $\mathbf{Q}_\ell^*$. Then, (76) imply the estimate

$$\frac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 \le (1+\beta)\mathbf{D}(\mathbf{\Lambda v},\mathbf{y}) + \left(1+\frac{1}{\beta}\right)\frac{1}{2}\mathbf{d}_\ell{}^2(\mathbf{y})$$

where $\mathbf{v}\in\mathbf{V_0}+\mathbf{u_0}$ and $\mathbf{y}\in\mathbf{Y}^*$. We rewrite this estimate as

$$\frac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 \le \mathcal{M}(\mathbf{v},\beta), \quad \forall \mathbf{v}\in\mathbf{V_0}+\mathbf{u_0}, \qquad \beta\in\mathbb{R}_+,$$

where

$$\mathcal{M}(\mathbf{v},\beta) := \inf_{\mathbf{y}\in\mathbf{Y}^*}\left\{(1+\beta)\mathbf{D}(\mathbf{\Lambda v},\mathbf{y}) + \left(1+\frac{1}{\beta}\right)\frac{1}{2}\mathbf{d}_\ell^2(\mathbf{y})\right\}.$$

Now, we proceed to finding computable upper bounds for the quantity $\mathbf{d}_\ell$. The first step is given by

Theorem

$$\tfrac{1}{2}\mathbf{d}_\ell^2(\mathbf{y}) \;=\; \sup_{\mathbf{w}\in\mathbf{V_0}} \left\{ -\tfrac{1}{2} \parallel\!\!| \mathbf{\Lambda w} |\!\!\parallel^2 - \langle \ell, \mathbf{w} \rangle - (\mathbf{y}, \mathbf{\Lambda w}) \right\}.$$

Proof

This assertion comes from that $\inf \mathcal{P} = \sup \mathcal{P}^*$. Indeed,

$$\frac{1}{2}\mathbf{d}_\ell^2(\mathbf{y}) = -\sup_{\boldsymbol{\eta}^* \in \mathbf{Q}_\ell^*} \left\{ -\frac{1}{2} \| \mathbf{y} - \boldsymbol{\eta}^* \|_*^2 \right\} = -\sup_{\boldsymbol{\eta}^* \in \mathbf{Q}_\ell^* - \mathbf{y}} \left\{ -\frac{1}{2} \| \boldsymbol{\eta}^* \|_*^2 \right\},$$

where $\mathbf{Q}_\ell^* - \mathbf{y} := \left\{ \boldsymbol{\eta}^* \in \mathbf{Y}^* \middle| \boldsymbol{\eta}^* = \text{æ}^* - \mathbf{y}, \qquad \text{æ}^* \in \mathbf{Q}_\ell^* \right\}.$
In other words, $\boldsymbol{\eta}^* \in \mathbf{Q}_\ell^* - \mathbf{y}$ if

$$(\boldsymbol{\eta}^*, \boldsymbol{\Lambda}\mathbf{w}) = -\langle \ell, \mathbf{w} \rangle - (\mathbf{y}, \boldsymbol{\Lambda}\mathbf{w}), \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

The right-hand side of this relation is a linear continuous functional. We denote it by ℓ_y and rewrite the relation as follows:

$$(\boldsymbol{\eta}^*, \boldsymbol{\Lambda}\mathbf{w}) + \langle \ell_\mathbf{y}, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

Then, $\mathbf{Q}^*_\ell - \mathbf{y} = \mathbf{Q}^*_{\ell_\mathbf{y}}$ and

$$\frac{1}{2}\mathbf{d}^2_\ell(\mathbf{y}) = - \sup_{\eta^* \in \mathbf{Q}^*_{\ell_\mathbf{y}}} \left\{ -\frac{1}{2} \| \eta^* \|^2_* \right\}.$$

This maximization problem is a form of **Problem** $\mathcal{P}^*$ if set $\mathbf{u_0} = \mathbf{0}$ and $\ell = \ell_y$. Since $\sup \mathcal{P}^* = \inf \mathcal{P}$, we have

$$\begin{aligned}
\frac{1}{2}\,\mathbf{d}^2_\ell(\mathbf{y}) &= - \inf_{w \in \mathbf{V_0}} \left\{ \frac{1}{2} \| \mathbf{\Lambda w} \|^2 + \langle \ell_\mathbf{y}, \mathbf{w} \rangle \right\} = \\
&= - \inf_{w \in \mathbf{V_0}} \left\{ \frac{1}{2} \| \mathbf{\Lambda w} \|^2 + \langle \ell, \mathbf{w} \rangle + (\mathbf{y}, \mathbf{\Lambda w}) \right\} = \\
&= \sup_{w \in \mathbf{V_0}} \left\{ -\frac{1}{2} \| \mathbf{\Lambda w} \|^2 - \langle \ell, \mathbf{w} \rangle - (\mathbf{y}, \mathbf{\Lambda w}) \right\}.
\end{aligned}$$

□

Corollary

We arrive at the conclusion that the majorant $\mathcal{M}(\mathbf{v}, \beta)$ has a minimax form

$$\mathcal{M}(\mathbf{v}, \beta) = \inf_{\mathbf{y} \in \mathbf{Y}^*} \sup_{\mathbf{w} \in \mathbf{V_0}} \left\{ (1+\beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \frac{\mathbf{1}+\beta}{\beta} \Big(-(\mathbf{y}, \mathbf{\Lambda w}) - \mathbf{J}(\mathbf{w}) \Big) \right\}. \quad (77)$$

Further,we use (77) for deriving upper and lower error bounds.

Upper estimates of $\| \mathbf{v} - \mathbf{u} \|$

In the relation

$$\mathcal{M}(\mathbf{v}, \beta) \le (1 + \beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \\ + \left(1 + \frac{1}{\beta} \right) \sup_{w \in \mathbf{V_0}} \left\{ -\frac{1}{2} \| \mathbf{\Lambda w} \|^2 - \langle \ell, \mathbf{w} \rangle - (\mathbf{y}, \mathbf{\Lambda w}) \right\},$$

we will estimate the value of supremum. Since $\mathbf{\Lambda}^*$ is the operator conjugate to $\mathbf{\Lambda}$, i.e.,

$$(\mathbf{y}, \mathbf{\Lambda w}) = \langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle, \qquad \forall \mathbf{w} \in \mathbf{V_0},$$

we have

$$\langle \ell, w \rangle + (\mathbf{y}, \mathbf{\Lambda w}) = \langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle \le \| \ell + \mathbf{\Lambda}^* \mathbf{y} \|_* \, \|\mathbf{\Lambda w}\|.$$

Here

$$\| \ell + \Lambda^* y \| := \sup_{w \in V_0} \frac{\langle \ell + \Lambda^* y, w \rangle}{\| \Lambda w \|} < +\infty.$$

We see that

$$\begin{aligned} & \sup_{w \in V_0} \left\{ -\tfrac{1}{2} \| \Lambda w \|^2 - \langle \ell, w \rangle - (y, \Lambda w) \right\} \leq \\ \leq \quad & \sup_{w \in V_0} \left\{ -\tfrac{1}{2} \| \Lambda w \|^2 + \| \ell + \Lambda^* y \| \, \|\Lambda w\| \right\} \leq \\ \leq \quad & \sup_{t > 0} \left\{ -\tfrac{1}{2} t^2 + \| \ell + \Lambda^* y \| \, t \right\} = \tfrac{1}{2} \| \ell + \Lambda^* y \|^2. \end{aligned}$$

Thus, we obtain

$$\frac{1}{2} \| \Lambda(\mathbf{v}-\mathbf{u}) \|^2 \leq (1+\beta)\mathbf{D}(\Lambda\mathbf{v},\mathbf{y}) + \frac{1+\beta}{2\beta} \| \ell + \Lambda^* \mathbf{y} \|^2. \qquad (78)$$

Deviation Majorant for the problem $\Lambda^* \mathcal{A} \Lambda u + \ell = 0$

$$(\mathcal{A}\Lambda(\mathbf{v}-\mathbf{u}), \Lambda(\mathbf{v}-\mathbf{u})) \leq$$

$$\leq (1+\beta)\Big((\mathcal{A}\Lambda\mathbf{v}, \Lambda\mathbf{v}) + (\mathcal{A}^{-1}\mathbf{y}, \mathbf{y}) - 2(\mathbf{y}, \Lambda\mathbf{v})\Big) +$$

$$+ \frac{1+\beta}{\beta} \mathbf{c}^2 \|\ell + \Lambda^* \mathbf{y}\|^2.$$

In the above, $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}, \quad \beta > 0, \quad \mathbf{y} \in \mathbf{U}.$

Theorem

For any $v \in \mathbf{V_0} + \mathbf{u_0}$,

$$\frac{1}{2} \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2 = \inf_{\substack{\mathbf{y} \in \mathbf{Q}^* \\ \beta > 0}} \mathbf{M}_\oplus(\mathbf{v}, \beta, \mathbf{y}).$$

If $\ell \in \mathbf{U}$, then $\mathbf{p} \in \mathbf{Q}^*$ and, therefore,

$$\inf_{\substack{\mathbf{y} \in \mathbf{Q}^* \\ \beta > 0}} \mathbf{M}_\oplus(\mathbf{v}, \beta, \mathbf{y}) \leq \mathbf{M}_\oplus(\mathbf{v}, \varepsilon, \mathbf{p}) = (1 + \varepsilon) \frac{1}{2} \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2,$$

where $\varepsilon > 0$ may be taken arbitrarily small.

Hence, the majorant $\mathbf{M}_\oplus$ is **reliable** and **exact**.

Lower estimates

Recall the minimax form of the Majorant

$$\mathcal{M}(\mathbf{v},\beta) = \inf_{\mathbf{y}\in\mathbf{Y}^*}\sup_{\mathbf{w}\in\mathbf{V_0}}\left\{(\mathbf{1}+\beta)\mathbf{D}(\mathbf{\Lambda v},\mathbf{y})+\frac{\mathbf{1}+\beta}{\beta}\Big(-(\mathbf{y},\mathbf{\Lambda w})-\mathbf{J}(\mathbf{w})\Big)\right\}.$$

Since $\sup\inf \le \inf\sup$, we have

$$\mathcal{M}(\mathbf{v},\beta) \ge \sup_{w\in\mathbf{V_0}}\inf_{\mathbf{y}\in\mathbf{Y}^*}\Big\{(1+\beta)D(\mathbf{\Lambda} v,\mathbf{y})- - \left(1+\frac{1}{\beta}\right)\left(\frac{\mathbf{1}}{\mathbf{2}}\,\|\!|\,\mathbf{\Lambda w}\,|\!\|^2+\langle\ell,\mathbf{w}\rangle+(\mathbf{y},\mathbf{\Lambda w})\right)\Big\}.$$

Thus, for any $\mathbf{w} \in \mathbf{V_0}$

$$\mathcal{M}(\mathbf{v},\beta) \geq \inf_{\mathbf{y}\in\mathbf{Y}^*}\left\{(1+\beta)\left(\frac{1}{2}\parallel \mathbf{y} \parallel_*^2 -(\mathbf{y},\mathbf{\Lambda v})\right)-\left(1+\frac{1}{\beta}\right)(\mathbf{y},\mathbf{\Lambda w})\right\}+ + (1+\beta)\frac{1}{2}\parallel \mathbf{\Lambda v} \parallel^2 - \left(1+\frac{1}{\beta}\right)\left(\frac{1}{2}\parallel \mathbf{\Lambda w} \parallel^2 +\langle \ell, \mathbf{w}\rangle\right),$$

Evidently, this estimate is also valid for the function $\beta\mathbf{w}$, which yields

$$\mathcal{M}(\mathbf{v},\beta) \geq (1+\beta)\inf_{\mathbf{y}\in\mathbf{Y}^*}\left\{\frac{1}{2}\parallel \mathbf{y} \parallel_*^2 -(\mathbf{y},\mathbf{\Lambda}(\mathbf{v}+\mathbf{w}))\right\}+ + (1+\beta)\left(\frac{1}{2}\parallel \mathbf{\Lambda v} \parallel^2 -\frac{\beta}{2}\parallel \mathbf{\Lambda w} \parallel^2 -\langle \ell, \mathbf{w}\rangle\right).$$

Note that

$$\inf_{\mathbf{y}\in\mathbf{Y}^*}\left\{\frac{\mathbf{1}}{\mathbf{2}}\,\|\!|\,\mathbf{y}\,|\!\|_*^2-(\,\mathbf{y},\mathbf{\Lambda}(\mathbf{v}+\mathbf{w})\,)\right\}\geq$$

$$\geq\inf_{\mathbf{y}\in\mathbf{Y}^*}\left\{\frac{\mathbf{1}}{\mathbf{2}}\,\|\!|\,\mathbf{y}\,|\!\|_*^2-\|\!|\,\mathbf{y}\,|\!\|_*\,\|\!|\,\mathbf{\Lambda}(\mathbf{v}+\mathbf{w})\,|\!\|\right\}=-\frac{\mathbf{1}}{\mathbf{2}}\,\|\!|\,\mathbf{\Lambda}(\mathbf{v}+\mathbf{w})\,|\!\|^2\,.$$

Thus, we obtain

$$\mathcal{M}(\mathbf{v},\beta)\geq(1+\beta)\Big\{-\frac{\mathbf{1}}{\mathbf{2}}\,\|\!|\,\mathbf{\Lambda}(\mathbf{v}+\mathbf{w})\,|\!\|^2+$$

$$+\frac{\mathbf{1}}{\mathbf{2}}\,\|\!|\,\mathbf{\Lambda}\mathbf{v}\,|\!\|^2-\frac{\beta}{\mathbf{2}}\,\|\!|\,\mathbf{\Lambda}\mathbf{w}\,|\!\|^2-\langle\ell,\mathbf{w}\rangle\Big\}=$$

$$=(1+\beta)\Big\{-(\mathcal{A}\mathbf{\Lambda}v,\mathbf{\Lambda}\mathbf{w})-\frac{\mathbf{1}+\beta}{\mathbf{2}}\,\|\!|\,\mathbf{\Lambda}\mathbf{w}\,|\!\|^2-\langle\ell,\mathbf{w}\rangle\Big\}.$$

In

$$(1+\beta)\Big\{-(\mathcal{A}\mathbf{\Lambda v},\mathbf{\Lambda w})-\frac{1+\beta}{2}\parallel\mathbf{\Lambda w}\parallel^2-\langle\ell,\mathbf{w}\rangle\Big\}.$$

$\mathbf{w}$ is an arbitrary function in $\mathbf{V_0}$. We may replace

$$\mathbf{w} \quad \text{by} \quad \frac{\mathbf{w}}{1+\beta}.$$

Such a replacement leads to the **Minorant $\mathbf{M_\ominus(v,w)}$** that gives a lower bound of the deviation from exact solution:

For any $\mathbf{w}\in\mathbf{V_0}$,

$$\frac{1}{2}\parallel\mathbf{\Lambda(v-u)}\parallel^2\geq-\frac{1}{2}\parallel\mathbf{\Lambda w}\parallel^2-(\mathcal{A}\mathbf{\Lambda v},\mathbf{\Lambda w})-\langle\ell,\mathbf{w}\rangle \tag{79}$$

Minorant is sharp

It is easy to see that

$$\sup_{\mathbf{w}\in\mathbf{V_0}} \mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w}) = \frac{\mathbf{1}}{\mathbf{2}} \parallel\!\!| \boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u}) |\!\!\parallel^{\mathbf{2}} .$$

Indeed, take $\mathbf{w} = \mathbf{u} - \mathbf{v}$.

$$\mathbf{M}_{\ominus}(\mathbf{v},\mathbf{u}-\mathbf{v}) = -\frac{\mathbf{1}}{\mathbf{2}} \parallel\!\!| \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) |\!\!\parallel^{\mathbf{2}} -(\mathcal{A}\boldsymbol{\Lambda}\mathbf{v}, \boldsymbol{\Lambda}\mathbf{u}-\mathbf{v}) - \langle \ell, \mathbf{u}-\mathbf{v}\rangle .$$

Represent the last two terms as follows:

$$\begin{aligned} &-(\mathcal{A}\boldsymbol{\Lambda}\mathbf{v}, \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) - \langle \ell, \mathbf{u}-\mathbf{v}\rangle = \\ &= -(\mathcal{A}\boldsymbol{\Lambda}\mathbf{v}, \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) + (\mathcal{A}\boldsymbol{\Lambda}\mathbf{u}, \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) = \\ &= (\mathcal{A}\boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}), \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) = \parallel\!\!| \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) |\!\!\parallel^{\mathbf{2}} \end{aligned}$$

so that this choice of $\mathbf{w}$ gives the true error.

Remark.

We outline that for the exact solution $\mathbf{M}_\ominus = \mathbf{M}_\oplus = \mathbf{0}$! Indeed, assume that $\mathbf{v}$ coincides with $\mathbf{u}$. In this case,

$$\mathbf{M}_\ominus(\mathbf{u},\mathbf{w}) = -\frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{\Lambda w} \|^2 - (\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda w}) - \langle \ell, \mathbf{w} \rangle = -\frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{\Lambda w} \|^2$$

and, therefore,

$$\sup_{\mathbf{w}\in\mathbf{V_0}} \mathbf{M}_\ominus(\mathbf{u},\mathbf{w}) = \mathbf{0}.$$

The same is true for the majorant. Indeed, set $\widehat{\mathbf{y}} = \mathcal{A}\mathbf{\Lambda u}$. Then,

$$\mathbf{M}_\oplus(\mathbf{u},\beta,\widehat{\mathbf{y}}) = (\mathbf{1}+\beta)\mathbf{D}(\mathbf{\Lambda u},\widehat{\mathbf{y}}) + \frac{\mathbf{1}+\beta}{\mathbf{2}\beta}\,\mathbf{c}^2 \|\ell + \mathbf{\Lambda}^* \mathcal{A}\mathbf{\Lambda u}\|^2 = \mathbf{0}.$$

Thus,

$$\inf_{\mathbf{y}\in\mathbf{Y}^*} \mathbf{M}_\oplus(\mathbf{u},\beta,\mathbf{y}) = \mathbf{0}.$$

Computability of two–sided estimates

By **computability** we mean that upper and lower estimates can be computed with any a priori given accuracy by solving finite-dimensional problems. In the case considered, they are certain problems for quadratic type integral functionals whose minimization (maximization) is performed by well-known methods.

Let $\{\mathbf{Y_i^*}\}_{\mathbf{i=1}}^{\infty}$ and $\{\mathbf{V_{0i}}\}_{\mathbf{i=1}}^{\infty}$ be two sequences of finite-dimensional subspaces that are dense in $\mathbf{Q}^*$ and $\mathbf{V_0}$, respectively, i.e., for any given $\boldsymbol{\varepsilon} > 0$ and arbitrary elements $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{w} \in \mathbf{V_0}$, one can find a natural number $\mathbf{k}_{\boldsymbol{\varepsilon}}$ such that

$$\inf_{\tilde{\mathbf{w}} \in \mathbf{V_{0i}}} \|\tilde{\mathbf{w}} - \mathbf{w}\|_{\mathbf{V}} \leq \boldsymbol{\varepsilon}, \qquad \inf_{\tilde{\mathbf{y}} \in \mathbf{Y_i^*}} \| \tilde{\mathbf{y}} - \mathbf{y} \|_{\mathbf{Q}^*} \leq \boldsymbol{\varepsilon}, \qquad \forall\, \mathbf{i} \geq \mathbf{k}_{\boldsymbol{\varepsilon}}.$$

Let us show that sequences of two-sided bounds converging to the actual error can be evaluated by minimizing the Majorant on $\{\mathbf{Y_i^*}\}$ and maximizing the Minorant on $\{\mathbf{V_{0i}}\}$.

Take a small $\varepsilon > 0$,. Then there exists a number $\mathbf{k}$ and elements $\mathbf{w_k} \in \mathbf{V_{0k}}$ and $\mathbf{p_k} \in \mathbf{Y^*_{0k}}$ satisfying the conditions

$$\|\mathbf{w_k} - (\mathbf{u} - \mathbf{v})\|_{\mathbf{V}} \leq \varepsilon, \qquad \| \mathbf{p_k} - \mathbf{p} \|_{\mathbf{Q}^*} \leq \varepsilon.$$

Define two quantities defined by solving **finite–dimensional problems**, namely

$$\mathbf{M}^{\mathbf{k}}_{\oplus} = \inf_{\substack{\mathbf{y_k} \in \mathbf{Y^*_k} \\ \beta \in \mathbb{R}_+}} \mathbf{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{y_k}), \quad \mathbf{M}^{\mathbf{k}}_{\ominus} = \sup_{\mathbf{w_k} \in \mathbf{V_{0k}}} \mathbf{M}_{\ominus}(\mathbf{v}, \mathbf{w_k}).$$

By the definition

$$\mathbf{M}_{\ominus}(\mathbf{v}, \mathbf{w_k}) \leq \mathbf{M}^{\mathbf{k}}_{\ominus} \leq \frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{u} - \mathbf{v} \|^2 \leq \mathbf{M}^{\mathbf{k}}_{\oplus} \leq \mathbf{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{p_k}).$$

The quantities $\mathbf{M}^{\mathbf{k}}_{\ominus}$ and $\mathbf{M}^{\mathbf{k}}_{\oplus}$ are **computable** (they require solving finite dimensional problems for quadratic type functionals). We will that

$$\mathbf{M}^{\mathbf{k}}_{\oplus} \rightarrow \frac{\mathbf{1}}{\mathbf{2}} \|\mathbf{\Lambda}(\mathbf{v}-\mathbf{u})\|^{\mathbf{2}},$$

$$\mathbf{M}^{\mathbf{k}}_{\ominus} \rightarrow \frac{\mathbf{1}}{\mathbf{2}} \|\mathbf{\Lambda}(\mathbf{v}-\mathbf{u})\|^{\mathbf{2}}$$

as the dimensionality $\mathbf{k}$ tends to $+\infty$.

Consider the **upper estimates**.

$$\mathsf{M}_{\oplus}(\mathbf{v},\beta,\mathbf{p_k}) = \quad (1+\beta)\mathsf{D}(\mathbf{\Lambda v},\mathbf{p_k}) + \frac{1+\beta}{2\beta}\mathbf{c}^2 \| \ell + \mathbf{\Lambda}^* \mathbf{p_k} \|^2 =$$

$$= (1+\beta)\mathsf{D}(\mathbf{\Lambda v},\mathbf{p_k}) + \frac{1+\beta}{2\beta}\mathbf{c}^2 \|\mathbf{\Lambda}^*(\mathbf{p_k}-\mathbf{p})\|^2.$$

Here

$$\mathsf{D}(\mathbf{\Lambda v},\mathbf{p_k}) = \frac{1}{2}(\mathbf{\Lambda v} - \mathcal{A}^{-1}\mathbf{p_k}, \mathcal{A}\mathbf{\Lambda v} - \mathbf{p_k}) =$$

$$= \frac{1}{2}\Big(\mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) - \mathcal{A}^{-1}(\mathbf{p_k}-\mathbf{p}), \mathcal{A}\mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) - (\mathbf{p_k}-\mathbf{p})\Big) =$$

$$= \frac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 + \| \mathbf{p_k}-\mathbf{p} \|_*^2 - (\mathbf{\Lambda}(\mathbf{v}-\mathbf{u}), \mathbf{p_k}-\mathbf{p}).$$

From the latter estimate we see that

$$\mathsf{D}(\mathbf{\Lambda v},\mathbf{p_k}) \leq \frac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 + \varepsilon \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \| + \frac{1}{2}\varepsilon^2. \qquad (80)$$

Since

$$\|\Lambda^*(p_k - p)\|_{Q^*} \le \varepsilon,$$

we find that

$$M^k_\oplus \le M_\oplus(v, \varepsilon, p_k) =$$
$$= (1+\varepsilon)\Big(\frac{1}{2}\,|||\,\Lambda(v-u)\,|||^2 + \varepsilon\,|||\,\Lambda(v-u)\,||| + \frac{1}{2}\varepsilon^2\Big) + \frac{1+\varepsilon}{2\varepsilon}c^2\varepsilon^2 =$$
$$= \frac{1}{2}\,|||\,\Lambda(v-u)\,|||^2 + c_4\varepsilon + o(\varepsilon^2).$$

where $c_4 = \frac{1}{2}\left(c + 2\,|||\,\Lambda(v-u)\,||| + |||\,\Lambda(v-u)\,|||^2\right)$. Thus, we conclude that

$$M^k_\oplus \longrightarrow \frac{1}{2}\,|||\,\Lambda(v-u)\,|||^2 \quad \text{as} \quad k \to \infty.$$

Remark

It is worth noting that the constant c_4 in the convergence term with ε depends on the norm of $(v - u)$, so that we can await that for a good approximation convergence of the upper bounds to the exact value of the error is faster than in the case where $\| v - u \|$ is considerable. This phenomenon was observed in many numerical experiments. In general, finding an upper bound for a precise approximation takes less CPU time than for a coarse one.

Consider the **lower estimates**.

$$\mathbf{M_\ominus(v, w_k)} = -\frac{1}{2} \| \mathbf{\Lambda w_k} \|^2 - (\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda w_k}) - \langle \ell, \mathbf{w_k} \rangle =$$

$$= -\frac{1}{2} \| \mathbf{\Lambda w_k} \| + (\mathcal{A}\mathbf{\Lambda}(\mathbf{u} - \mathbf{v}), \mathbf{\Lambda w_k}) =$$

$$= \frac{1}{2} \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2 - \frac{1}{2} \| \mathbf{\Lambda}(\mathbf{w_k} - (\mathbf{u} - \mathbf{v})) \|^2 \geq$$

$$\geq \frac{1}{2} \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2 - \frac{1}{2} \mathbf{c_2} \|\mathbf{\Lambda}(\mathbf{w_k} - (\mathbf{u} - \mathbf{v}))\|^2 .$$

This implies the estimate

$$\frac{1}{2} \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2 \geq \mathbf{M_\ominus^k} \geq \frac{1}{2} \|\mathbf{\Lambda}(\mathbf{u} - \mathbf{v})\|^2 - \mathbf{c_5} \boldsymbol{\varepsilon}^2,$$

where $\mathbf{c_5} > \mathbf{0}$ depends on the norm of $\mathbf{\Lambda}$. Thus,

$$\mathbf{M_\ominus^k} \to \frac{1}{2} \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^2 \quad \text{as } \mathbf{k} \to \infty.$$

Computable upper bound of the effectivity index

Having $\mathbf{M}^{\mathbf{k}}_{\oplus}$ and $\mathbf{M}^{\mathbf{k}}_{\ominus}$, one can define the number

$$\eta_{\mathbf{k}} := \frac{\mathbf{M}^{\mathbf{k}}_{\oplus}}{\mathbf{M}^{\mathbf{k}}_{\ominus}} \geq \mathbf{1}, \tag{81}$$

which gives an idea of the **quality of the error estimation**. From the above it follows that

$$\eta_{\mathbf{k}} \to \mathbf{1}, \qquad \text{as } \mathbf{k} \to +\infty.$$

Relationships with other methods

$$\mathbf{M}_{\oplus}(\mathbf{v},\beta,\mathbf{y}) = (\mathbf{1}+\beta)\mathbf{D}(\mathbf{\Lambda v},\mathbf{y}) + \frac{\mathbf{1}+\beta}{2\beta}\mathbf{c}^2 \| \boldsymbol{\ell} + \mathbf{\Lambda}^*\mathbf{y} \|^2$$

involves an arbitrary function $\mathbf{y}$. We are aimed to show that some special choices of it lead to known error estimates.

We assume that $\langle \boldsymbol{\ell}, \mathbf{w} \rangle = (\mathbf{g}, \mathbf{w})$, where $\mathbf{g} \in \mathbf{U}$, so that $\mathbf{p} \in \mathbf{Q}^* \subset \mathbf{Q}^*_{\boldsymbol{\ell}}$ and

$$\mathbf{Q}^*_{\boldsymbol{\ell}} := \{\mathbf{y} \in \mathbf{Q}^* \mid (\mathbf{\Lambda}^*\mathbf{y} + \mathbf{g}, \mathbf{w}) = \mathbf{0}, \qquad \forall \mathbf{w} \in \mathbf{V_0}\}.$$

Let us first define the function

$$\mathbf{y_0} = \mathcal{A}\mathbf{\Lambda v}. \tag{82}$$

A variety of options comes from the relation

$$\mathbf{y} = \mathbf{\Pi y_0}, \tag{83}$$

where $\mathbf{\Pi}$ is a certain continuous mapping.

Residual based estimate

If $\mathbf{\Pi}$ is the identity mapping of $\mathbf{Y}^*$, i.e., $\mathbf{y} = \mathbf{y_0} :== \mathcal{A}\mathbf{\Lambda v}$, then

$$\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}_0^*) = \mathbf{0}.$$

Use the majorant in the form (78):

$$\frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^2 \leq (\mathbf{1} + \beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \frac{\mathbf{1} + \beta}{\mathbf{2}\beta} \| \ell + \mathbf{\Lambda}^* \mathbf{y} \|^2.$$

Now, it contains only the second term, which after the minimization with respect to β gives

$$\| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \| \leq \| \ell + \mathbf{\Lambda}^* \mathcal{A} \mathbf{\Lambda v} \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{(\ell, \mathbf{w}) + (\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda w})}{\| \mathbf{\Lambda w} \|}. \quad (84)$$

If $\mathbf{v}$ is obtained by FEM and $\mathbf{v} = \mathbf{u_h} \in \mathbf{V_h} := \mathbf{V_{0h}} + \mathbf{u_0}$, then we arrive at the following estimate:

$$\| \mathbf{\Lambda}(\mathbf{u_h} - \mathbf{u}) \| \leq \mathbf{I}\, \ell + \mathbf{\Lambda}^* \mathcal{A} \mathbf{\Lambda} \mathbf{u_h} \,\mathbf{I} = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{(\ell, \mathbf{w} - \mathbf{w_h}) + (\mathcal{A}\mathbf{\Lambda}\mathbf{u_h}, \mathbf{\Lambda}(\mathbf{w} - \mathbf{w_h}))}{\| \mathbf{\Lambda} \mathbf{w} \|}.$$

We find an upper bound of the right–hand side by the arguments accepted in the classical residual method.

Conclusion

If in the functional a posteriori error estimate is applied to a FEM solution $\mathbf{u_h}$ then we may select the variable $\mathbf{y}$ in the simplest way as

$$\mathbf{y} = \mathbf{\Lambda u_h}.$$

Then, if $\mathbf{u_h}$ is a Galerkin approximation, we can use this fact and obtain at an upper bound given by the *residual type a posteriori error estimate* that involves integral terms associated with finite elements and interelement jumps.

Estimates using post–processing of the dual variable

In $\mathbf{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{y})$ the best choice is $\mathbf{y} = \mathbf{p} \in \mathbf{Q}^*$. Therefore, if $\mathbf{y_0} \notin \mathbf{Q}^*$ then its mapping $\mathbf{Q}^*$ could be a better approximation of $\mathbf{p}$. Let us denote such a mapping by $\mathbf{\Pi_1}$. We obtain

$$\mathbf{y_1} = \mathbf{\Pi_1 y_0} \in \mathbf{Q}^* \tag{85}$$

and the quantity $\mathbf{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{y_1})$, which leads to the error majorant

$$\mathbf{M}^{(1)}_{\oplus}(\mathbf{v}) = \inf_{\beta \in \mathbb{R}_+} \Big\{ (1+\beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{\Pi}_1(\mathcal{A}\mathbf{\Lambda v})) + \\ + \frac{1+\beta}{2\beta} \mathbf{c}^2 \, \| \, \ell + \mathbf{\Lambda}^* \mathbf{\Pi}_1(\mathcal{A}\mathbf{\Lambda v}) \, \|^2 \Big\}. \tag{86}$$

Particular case

In the simplest case associated with the problem

$$\mathbf{\Delta u} + \mathbf{f} = \mathbf{0}, \qquad \mathbf{u} = \mathbf{u_0} \quad \text{on } \partial\mathbf{\Omega}$$

we have

$$\mathbf{M}^{(1)}_{\oplus}(\mathbf{u_h}) =$$

$$= \inf_{\beta\in\mathbb{R}_+} \left\{ (\mathbf{1}+\beta)\|\nabla\mathbf{u_h} - \mathbf{\Pi_1}(\nabla\mathbf{u_h})\|^2 + \frac{(\mathbf{1}+\beta)\mathbf{C}^2_{\Omega}}{\mathbf{2}\beta}\|\mathbf{f} + \mathbf{div}\mathbf{\Pi_1}(\nabla\mathbf{u_h})\|^2 \right\}.$$

If $\mathbf{\Pi_1}$ is a gradient averaging operator, then the first term in the right–hand side is **the difference between the original and averaged gradient**, i.e. it coincides with a **gradient averaging indicator**. However, as we have seen in previous lectures, such an indicator cannot provide a reliable upper bound of the error. The second term in the right-hand side shows what is necessary to add in order to provide the reliability.

Estimates based on the "equilibration" of the dual variable

Let $\mathbf{\Pi_2}$ maps $\mathbf{Y}^*$ to the set $\mathbf{Q}^*_\ell$. Define

$$\mathbf{y_2} = \mathbf{\Pi_2 y_0} \in \mathbf{Q}^*_\ell. \tag{87}$$

Then,

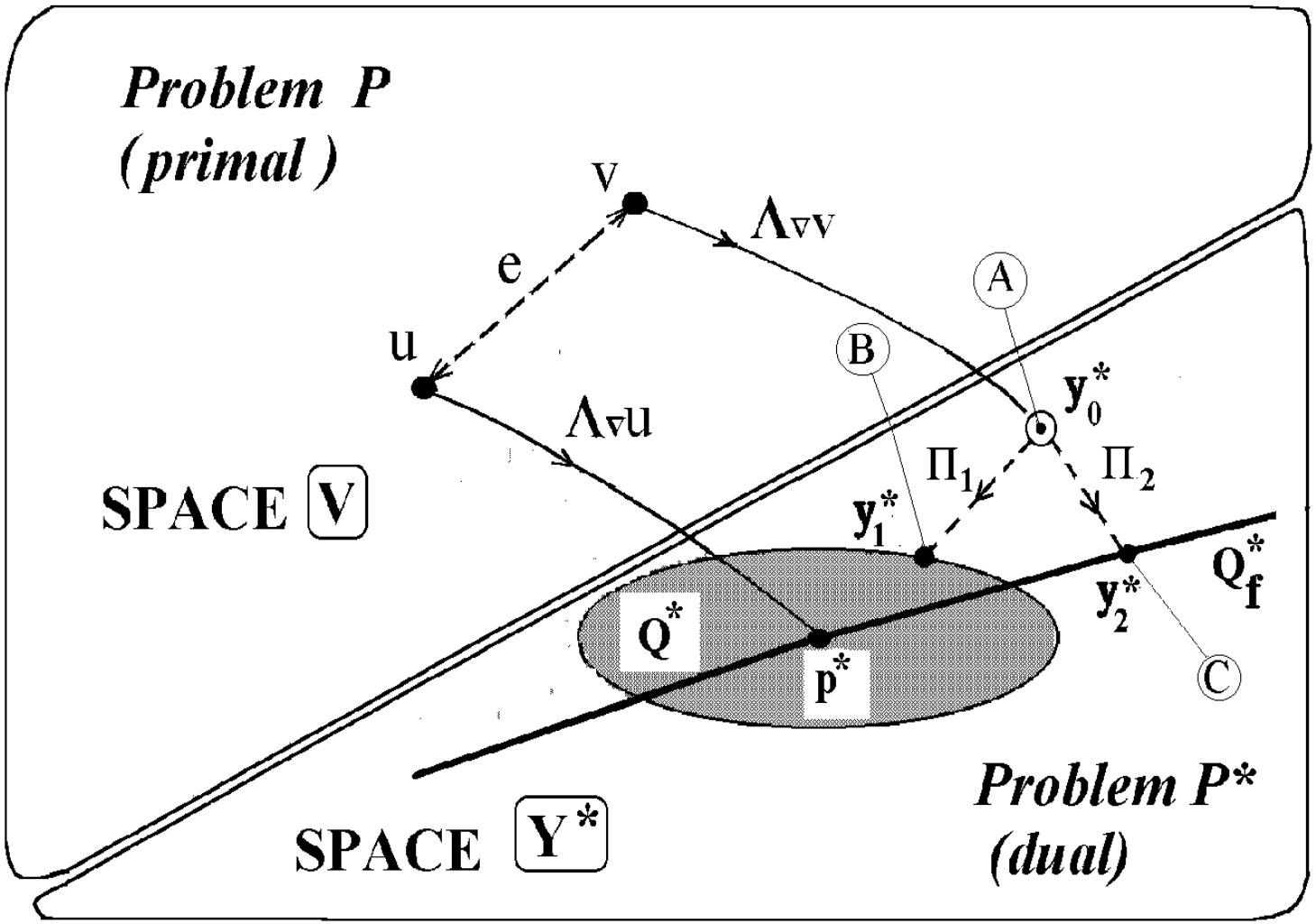
$$\mathbf{\Lambda}^* \mathbf{y_2} + \ell = \mathbf{0},$$

so that the Majorant has only the first term:

$$\mathbf{M}^{(2)}_{\oplus}(\mathbf{v}) = \mathbf{D}(\mathbf{\Lambda v}, \mathbf{y_2}).$$

$\mathbf{\Pi_2}$ is natural to call an **equilibration operator**. In general, it is rather difficult to construct an "exact mapping" $\mathbf{\Pi_2}$ to $\mathbf{Q}^*_\ell$. One may use an operator $\widetilde{\mathbf{\Pi}}_\mathbf{2}$, which provides an approximate "equilibration". In this case, the **second term of the Majorant does not vanish and should be taken into account.**

Various choices of the dual variable lead to certain a posteriori methods



A priori projection type error estimates

As an exercise, we now will derive classical a priori projection type error estimates from a functional a posteriori estimate. Let $\mathbf{u_h} \in \mathbf{V_h}$ be a Galerkin approximation of $\mathbf{u}$. We have

$$\| \mathbf{\Lambda}(\mathbf{u}-\mathbf{u_h}) \|^2 \leq \mathbf{2}(\mathbf{1}+\beta)\mathbf{D}(\mathbf{\Lambda u_h}, \mathbf{y}) + \left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right) \| \mathbf{\Lambda}^* \mathbf{y} + \ell \|^2$$

Set here $\mathbf{y} = \mathcal{A}\mathbf{\Lambda v_h}$, where $\mathbf{v_h}$ is an arbitrary element of $\mathbf{V_h}$. Then,

$$\| \mathbf{\Lambda}^* \mathbf{y} + \ell \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{(\mathbf{y} - \mathbf{p}, \mathbf{\Lambda w})}{\| \mathbf{\Lambda w} \|} =$$

$$= \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{(\mathcal{A}\mathbf{\Lambda}(\mathbf{v_h} - \mathbf{u}), \mathbf{\Lambda w})}{\| \mathbf{\Lambda w} \|} \leq \| \mathbf{\Lambda}(\mathbf{v_h} - \mathbf{u}) \| .$$

It is easy to see that

$$\mathsf{D}(\mathbf{\Lambda u_h}, \mathcal{A}\mathbf{\Lambda v_h}) = \mathsf{J}(\mathbf{v_h}) - \mathsf{J}(\mathbf{u_h}).$$

Indeed,

$$\begin{aligned}\mathsf{D}(\mathbf{\Lambda u_h}, \mathcal{A}\mathbf{\Lambda v_h}) = \frac{1}{2}(\mathcal{A}\mathbf{\Lambda v_h}, \mathbf{\Lambda v_h}) + \langle \ell, \mathbf{v_h}\rangle - \\ - \frac{1}{2}(\mathcal{A}\mathbf{\Lambda u_h}, \mathbf{\Lambda u_h}) - \langle \ell, \mathbf{u_h}\rangle + \\ + (\mathcal{A}\mathbf{\Lambda u_h}, \mathbf{\Lambda}(\mathbf{u_h} - \mathbf{v_h})) + \langle \ell, \mathbf{u_h} - \mathbf{v_h}\rangle.\end{aligned}$$

Since $\mathbf{u_h} \in \mathbf{V_h}$ is a Galerkin approximation, the last two terms vanish and we obtain the relation.
We know that

$$\begin{aligned}\| \mathbf{\Lambda}(\mathbf{u_h} - \mathbf{u}) \|^2 &= 2(\mathsf{J}(\mathbf{u_h}) - \mathsf{J}(\mathbf{u})), \\ \| \mathbf{\Lambda}(\mathbf{v_h} - \mathbf{u}) \|^2 &= 2(\mathsf{J}(\mathbf{v_h}) - \mathsf{J}(\mathbf{u})).\end{aligned}$$

Therefore,

$$\mathbf{2D}(\mathbf{\Lambda u_h}, \mathcal{A}\mathbf{\Lambda v_h}) = \mathbf{2(J(v_h) - J(u))} - \mathbf{2(J(u_h) - J(u))} = \\ = \| \mathbf{\Lambda(v_h - u)} \|^2 - \| \mathbf{\Lambda(u_h - u)} \|^2 .$$

Now, the error estimate comes in the form

$$\| \mathbf{\Lambda(u - u_h)} \|^2 \leq (\mathbf{1} + \beta)(\| \mathbf{\Lambda(v_h - u)} \|^2 - \| \mathbf{\Lambda(u_h - u)} \|^2) + \\ + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \| \mathbf{\Lambda(v_h - u)} \|^2 .$$

Thus, we obtain

$$(\mathbf{2} + \beta) \| \mathbf{\Lambda(u - u_h)} \|^2 \leq \\ \leq (\mathbf{1} + \beta) \| \mathbf{\Lambda(v_h - u)} \|^2 + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \| \mathbf{\Lambda(v_h - u)} \|^2,$$

We see that

$$\| \mathbf{\Lambda}(\mathbf{u} - \mathbf{u_h}) \|^2 \leq \left(1 + \frac{1}{\beta(2+\beta)} \right) \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v_h}) \|^2 .$$

Since β is an arbitrary positive number, we arrive at the projection type error estimate

$$\| \mathbf{\Lambda}(\mathbf{u} - \mathbf{u_h}) \| \leq \inf_{\mathbf{v_h} \in \mathbf{V_h}} \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v_h}) \| .$$

APPLICATIONS TO PARTICULAR CLASSES OF PARTIAL DIFFERENTIAL EQUATIONS

Diffusion equation

Let $\mathcal{A}$ is produced by a matrix $\mathbf{A} = \{\mathbf{a_{ij}}\} = \{\mathbf{a_{ji}}\}$, $\mathbf{V} = \mathbf{H^1(\Omega)}$, where Ω is a Lipschitz domain, $\mathbf{U} = \mathbf{L^2(\Omega, \mathbb{R}^n)}$, and $\mathbf{\Lambda w} = \nabla \mathbf{w}$. Let the entries of $\mathbf{A}$ be bounded at almost all points of Ω and such that

$$\mathbf{c_1}|\xi|^2 \leq \mathbf{a_{ij}}\xi_\mathbf{i}\xi_\mathbf{j} \leq \mathbf{c_2}|\xi|^2, \qquad \forall \xi \in \mathbb{R}^\mathbf{n}. \tag{88}$$

Then, the spaces $\mathbf{Y}$ and $\mathbf{Y}^*$ have the norms

$$\| \mathbf{y} \|^2 = \int_\Omega \mathbf{A}\mathbf{y} \cdot \mathbf{y}\, \mathbf{dx}, \qquad \| \mathbf{y} \|_*^2 = \int_\Omega \mathbf{A}^{-1}\mathbf{y} \cdot \mathbf{y}\, \mathbf{dx}.$$

Dirichlet boundary conditions

We begin with the problem

$$\mathbf{div A}\nabla \mathbf{u} = \mathbf{f} \quad \text{in} \quad \Omega, \tag{89}$$

$$\mathbf{u} = \mathbf{u_0} \quad \text{on} \quad \partial\Omega. \tag{90}$$

In this case, $\mathbf{V_0} = \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\Omega)$ and $\mathbf{u}$ meets the integral identity

$$\int_\Omega \mathbf{A}\nabla \mathbf{u} \cdot \nabla \mathbf{w}\, \mathbf{dx} + \langle \mathbf{f}, \mathbf{w} \rangle = \mathbf{0}, \qquad \forall \mathbf{w} \in \mathbf{V_0}. \tag{91}$$

The relation $(\mathbf{y}, \mathbf{\Lambda w}) = \langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle$ has the form

$$\int_\Omega \mathbf{y} \cdot \nabla \mathbf{w}\, \mathbf{dx} = \langle -\mathbf{div}\, \mathbf{y}, \mathbf{w} \rangle,$$

where $\mathbf{\Lambda}^* = -\mathbf{div}$ and $\mathbf{div}\, \mathbf{y}$ is in $\mathbf{H}^{-\mathbf{1}}(\Omega)$.

The operator $\mathbf{\Lambda}$ satisfies the required inequality

$$\mathbf{c_\Omega}\|\nabla \mathbf{w}\| \geq \|\mathbf{w}\|, \qquad \forall \mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\mathbf{\Omega}).$$

Upper estimates of $\|\!|\!| \mathbf{v} - \mathbf{u} |\!|\!|$ for an approximation $v \in V_0 + u_0$ follow from the general estimate presented in Lecture 5. We have

$$\frac{\mathbf{1}}{\mathbf{2}} \int_{\mathbf{\Omega}} \mathbf{A}\nabla(\mathbf{v}-\mathbf{u}) \cdot \nabla(\mathbf{v}-\mathbf{u})\, \mathbf{dx} \leq \mathbf{M}_{\oplus}(\mathbf{v}, \boldsymbol{\beta}, \mathbf{y}),$$

where

$$\mathbf{M}_{\oplus}(\mathbf{v}, \boldsymbol{\beta}, \mathbf{y}) = \frac{\mathbf{1}+\boldsymbol{\beta}}{\mathbf{2}} \int_{\mathbf{\Omega}} \left(\nabla \mathbf{v} - \mathbf{A}^{-1}\mathbf{y} \right) \cdot \left(\mathbf{A}\nabla \mathbf{v} - \mathbf{y} \right) \mathbf{dx} + \frac{\mathbf{1}+\boldsymbol{\beta}}{\mathbf{2}\boldsymbol{\beta}} \frac{\mathbf{c}_{\mathbf{\Omega}}^{\mathbf{2}}}{\mathbf{c}_{\mathbf{1}}^{\mathbf{2}}} \|\mathbf{div}\, \mathbf{y} - \mathbf{f}\|^{\mathbf{2}} \quad (92)$$

Certainly, the above estimate is aplicable for the case $\mathbf{f} \in \mathbf{L}^2(\Omega)$ so that

$$\langle \mathbf{f}, \mathbf{w} \rangle = \int_{\Omega} \mathbf{f}\mathbf{w}\, \mathbf{dx},$$

and for $\mathbf{y} \in \mathbf{H}(\Omega, \mathbf{div})$.
Let $\{\mathbf{Y}^*_{\mathbf{k}}\}$ be finite-dimensional subspaces of $\mathbf{Y}^*$ such that

$$\mathbf{Y}^*_{\mathbf{k}} \in \mathbf{H}(\Omega, \mathbf{div}) \quad \text{for all } \mathbf{k} = \mathbf{1}, \mathbf{2}, ...;$$
$$\dim \mathbf{Y}^*_{\mathbf{k}} \to +\infty \quad \text{as } \mathbf{k} \to \infty.$$

We obtain computable upper bounds

$$\mathbf{M}^{\mathbf{k}}_{\oplus} = \inf_{\substack{\mathbf{y} \in \mathbf{Y}^*_{\mathbf{k}} \\ \beta \in \mathbb{R}_+}} \Big\{ \frac{\mathbf{1}+\beta}{\mathbf{2}} \int_{\Omega} (\nabla \mathbf{v} - \mathbf{A}^{-1}\mathbf{y}) \cdot (\mathbf{A}\nabla \mathbf{v} - \mathbf{y})\, \mathbf{dx} + \\ + \frac{1+\beta}{2\beta} \frac{\mathbf{c}^2_{\Omega}}{\mathbf{c_1}} \| \mathbf{div}\, \mathbf{y} - \mathbf{f} \|^2_{\Omega} \Big\}. \quad (93)$$

Lower estimates

We have

$$\frac{1}{2}\int_{\Omega} \mathbf{A}\nabla(\mathbf{v}-\mathbf{u})\cdot\nabla(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \geq \mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w}), \qquad \forall \mathbf{w}\in\mathbf{V_0},$$

where

$$\mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w}) = -\frac{\mathbf{1}}{\mathbf{2}}\int_{\Omega}\mathbf{A}\nabla\mathbf{w}\cdot\nabla\mathbf{w}\,\mathbf{dx} - \int_{\Omega}\mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{w}\mathbf{dx} - \langle\mathbf{f},\mathbf{w}\rangle.$$

Let $\{\mathbf{V_{0k}}\}$ be finite-dimensional subspaces such that

$$\mathbf{V}^{*}_{\mathbf{0k}} \in \mathbf{V_0} \quad \text{for all } \mathbf{k}=\mathbf{1},\mathbf{2},...;$$
$$\dim\mathbf{V_{0k}} \to +\infty \quad \text{as } \mathbf{k}\to\infty.$$

Find the numbers

$$\mathbf{M}^{\mathbf{k}}_{\ominus} = \sup_{\mathbf{w_k}\in\mathbf{V_{0k}}} \mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w_k}). \tag{94}$$

Both sequences $\mathbf{M}^{\mathbf{k}}_{\ominus}$ and $\mathbf{M}^{\mathbf{k}}_{\oplus}$ tend to $\frac{1}{2}\,|\!|\!|\,\mathbf{v}-\mathbf{u}\,|\!|\!|^2$ as $\mathbf{k}\to\infty$, provided that $\{\mathbf{Y}^*_{\mathbf{k}}\}$ and $\{\mathbf{V_{0k}}\}$ possess necessary approximation properties (limit density). Note that if $\mathbf{v}$ is a Galerkin approximation computed on $\mathbf{V_{0k}}$, then $\mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w_k})=\mathbf{0}$. This means that to obtain a sensible lower estimate in this case, one must always use a finite-dimensional subspace that is *larger* than $\mathbf{V_{0k}}$.

Neumann boundary condition

Consider the Neumann boundary condition

$$\boldsymbol{\nu} \cdot \mathbf{A}\nabla \mathbf{u} + \mathbf{F} = \mathbf{0} \qquad \text{on} \qquad \partial\Omega, \tag{95}$$

where $\boldsymbol{\nu}$ is the vector of unit outward normal to $\partial\Omega$. To apply the general scheme we set

$$\mathbf{V_0} := \left\{ \mathbf{v} \in \mathbf{H}^1(\Omega) \;\middle|\; \int_\Omega \mathbf{v}\,\mathbf{dx} = \mathbf{0} \right\}$$

and define $\boldsymbol{\Lambda}^* \mathbf{y} \in \mathbf{V_0^*}$ by the relation

$$\langle \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_\Omega \mathbf{y} \cdot \nabla \mathbf{w}\,\mathbf{dx}, \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

If $\mathbf{y}$ is sufficiently regular then

$$\langle \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_\Omega (-\mathbf{div y})\mathbf{w}\,\mathbf{dx} + \int_{\partial\Omega} (\mathbf{y} \cdot \boldsymbol{\nu})\mathbf{w}\mathbf{dx}.$$

Therefore, in such a case

$$\mathbf{\Lambda}^* \mathbf{y} = [-\mathbf{div}\, \mathbf{y} \mid_{\Omega}; \qquad \mathbf{y} \cdot \boldsymbol{\nu} \mid_{\partial\Omega}]$$

Also, we assume that $\mathbf{F}$ and $\mathbf{f}$ satisfy the equilibrium condition

$$\int_{\Omega} \mathbf{f}\, \mathbf{dx} + \int_{\partial\Omega} \mathbf{F}\, \mathbf{dx} = \mathbf{0}.$$

Assume that $\mathbf{f} \in \mathbf{L^2(\Omega)}$ and $\mathbf{F} \in \mathbf{L^2}(\partial\mathbf{\Omega})$. Then the Neumann problem has a solution defined by the integral identity

$$\int_{\Omega} \mathbf{A}\nabla\mathbf{u} \cdot \nabla\mathbf{w}\, \mathbf{dx} + \langle \ell, \mathbf{w} \rangle = \mathbf{0}, \qquad \forall \mathbf{w} \in \mathbf{V_0},$$

where

$$\langle \ell, \mathbf{w} \rangle = \int_{\Omega} \mathbf{f}\mathbf{w}\, \mathbf{dx} + \int_{\partial\Omega} \mathbf{F}\mathbf{w}\, \mathbf{ds}.$$

In general, $\mathbf{]}\ell + \mathbf{\Lambda}^* \mathbf{y}\mathbf{[}$ is estimated in terms of the norms

$$\| \operatorname{\mathbf{div}} \mathbf{y} - \mathbf{f} \|_{\mathbf{H}^{-1}} \quad \text{and} \quad \| \mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F} \|_{\mathbf{H}^{-1/2}}.$$

However, if we assume that $\mathbf{y}$ possesses a certain regularity, so that

$$\mathbf{y} \in \mathbf{Q}^*(\Omega) := \{\mathbf{y} \in \mathbf{Y}^* \mid \operatorname{\mathbf{div}} \mathbf{y} \in \mathbf{L}^2(\Omega), \mathbf{y} \cdot \boldsymbol{\nu} \in \mathbf{L}^2(\partial\Omega)\},$$

then

$$\langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_{\Omega} (\mathbf{f} - \operatorname{\mathbf{div}} \mathbf{y}) \mathbf{w} \, \mathbf{dx} + \int_{\partial\Omega} (\mathbf{F} + \mathbf{y} \cdot \boldsymbol{\nu}) \mathbf{w} \, \mathbf{ds}$$

and, therefore,

$$|\langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle| \leq \\ \leq \| \operatorname{\mathbf{div}} \mathbf{y} - \mathbf{f} \|_{2,\Omega} \|\mathbf{w}\|_{2,\Omega} + \|\mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F}\|_{2,\partial\Omega} \|\mathbf{w}\|_{2,\partial\Omega}. \quad (96)$$

Let the constant $\mathbf{c_\Omega}$ be defined as

$$\frac{1}{\mathbf{c}^2_{(\Omega,\partial\Omega)}} = \inf_{\mathbf{w}\in\mathbf{V_0}} \frac{\int_\Omega \mathbf{A}\nabla\mathbf{w}\cdot\nabla\mathbf{w}\,\mathbf{dx}}{\|\mathbf{w}\|^2_{2,\Omega} + \|\mathbf{w}\|^2_{2,\partial\Omega}}.$$

Since the trace operator is bounded, this constant is finite. Therefore, (96) implies the estimate

$$|\langle \ell + \mathbf{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle| \le \mathbf{c}_{(\Omega,\partial\Omega)} \left(\| \operatorname{\mathbf{div}} \mathbf{y} - \mathbf{f} \|^2_{2,\Omega} + \|\mathbf{y}\cdot\nu + \mathbf{F}\|^2_{2,\partial\Omega} \right)^{1/2} \| \mathbf{\Lambda w} \|^2$$

and the second term of the majorant is calculated as follows:

$$\| \ell + \mathbf{\Lambda}^*\mathbf{y} \| = \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\langle \ell + \mathbf{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle}{\| \mathbf{\Lambda w} \|} \le \mathbf{c}_{(\Omega,\partial\Omega)} \left(\| \operatorname{\mathbf{div}} \mathbf{y} - \mathbf{f} \|^2_{2,\Omega} + \|\mathbf{y}\cdot\nu + \mathbf{F}\|^2_{2,\partial\Omega} \right)^{1/2}.$$

The term $\mathbf{D}(\mathbf{\Lambda}\nu, \mathbf{y})$ is defined as in the Dirichlét problem.

We see that the Majorants $\mathbf{M}_\oplus$ for the two main boundary-value problems have different values of $\mathbf{c_\Omega}$. In addition, the Neumann problem majorant contains an extra term

$$\|\mathbf{y}\cdot\boldsymbol{\nu}+\mathbf{F}\|_{2,\partial\Omega}$$

that penalizes violations of the Neumann boundary condition.
It is worth noting that if the given $\mathbf{F}$ can be exactly reproduced by $\mathbf{y}\cdot\boldsymbol{\nu}$ for $\mathbf{y}$ in a certain finite dimensional subspace $\mathbf{Y_k^*}$, then one can compute $\mathbf{M}_\oplus^{\mathbf{k}}$ as

$$\mathbf{M}_\oplus^{\mathbf{k}} = \inf_{\substack{\mathbf{y}\in\mathbf{Y_k^*},\ \mathbf{y}\cdot\nu=\mathbf{F}\ \text{on}\ \partial\Omega \\ \beta\in\mathbb{R}_+}} \Big\{ \frac{1+\beta}{2}\int_\Omega (\nabla\mathbf{v}-\mathbf{A}^{-1}\mathbf{y})\cdot(\mathbf{A}\nabla\mathbf{v}-\mathbf{y})\,\mathbf{dx} + \frac{1+\beta}{2\beta}\,\mathbf{c}^2_{(\Omega,\partial\Omega)}\|\,\mathbf{div}\,\mathbf{y}-\mathbf{f}\,\|^2_\Omega\Big\}. \quad (97)$$

Mixed boundary conditions

Let $\partial\Omega$ consist of two measurable nonintersecting parts $\partial_1\Omega$ and $\partial_2\Omega$, on which different boundary conditions are given:

$$\mathbf{u} = \mathbf{u_0} \quad \text{on} \quad \partial_1\mathbf{\Omega},$$

$$\boldsymbol{\nu} \cdot \mathbf{A}\nabla\mathbf{u} + \mathbf{F} = \mathbf{0} \quad \text{on} \quad \partial_2\mathbf{\Omega}.$$

Set

$$\mathbf{V_0} := \left\{ \mathbf{v} \in \mathbf{H}^1(\mathbf{\Omega}) \,|\, \mathbf{v} = \mathbf{0} \qquad \text{on} \qquad \partial_1\mathbf{\Omega} \right\}$$

and

$$\langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_{\mathbf{\Omega}} \mathbf{y} \cdot \nabla \mathbf{w} \, \mathbf{dx}, \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

Assume that

$$\mathbf{f} \in \mathbf{L}^2(\Omega), \qquad \mathbf{F} \in \mathbf{L}^2(\partial_2\Omega).$$

and $\mathbf{y}$ possesses an extra regularity, namely,

$$\mathbf{y} \in \mathbf{Q}^*(\Omega) := \left\{ \mathbf{y} \in \mathbf{Y}^* \,|\, \mathbf{div}\, \mathbf{y} \in \mathbf{L}^2(\Omega),\ \mathbf{y}\cdot\boldsymbol{\nu} \in \mathbf{L}^2(\partial_2\Omega) \right\}.$$

Then, for any $\mathbf{w} \subset \mathbf{V_0}$, we have

$$\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_\Omega (\mathbf{div}\, \mathbf{y} - \mathbf{f}) \mathbf{w}\, \mathbf{dx} + \int_{\partial_2\Omega} (\mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F}) \mathbf{w}\, \mathbf{ds},$$

Note that $\mathbf{p} \in \mathbf{Q}^*(\Omega)$!

Now, we obtain

$$|\langle \ell + \Lambda^* \mathbf{y}, \mathbf{w} \rangle| \le \|\mathbf{div}\mathbf{y} - \mathbf{f}\|_{2,\Omega} \|\mathbf{w}\|_{2,\Omega} + \\ + \|\mathbf{y} \cdot \nu + \mathbf{F}\|_{2,\partial_2\Omega} \|\mathbf{w}\|_{2,\partial_2\Omega}.$$

Let γ and γ_* be two numbers such that $\gamma > 1, \qquad \gamma_* > 1, \ \frac{1}{\gamma} + \frac{1}{\gamma_*} = 1.$
Use the algebraic inequality

$$ab + cd \le \sqrt{\gamma a^2 + \gamma_* c^2} \sqrt{\frac{1}{\gamma} b^2 + \frac{1}{\gamma_*} d^2}.$$

Then

$$|\langle \ell + \Lambda^* \mathbf{y}, \mathbf{w} \rangle| \le \left(\gamma \|\mathbf{div}\, \mathbf{y} - \mathbf{f}\|^2_{2,\Omega} + \gamma_* \|\mathbf{y} \cdot \nu + \mathbf{F}\|^2_{2,\partial_2\Omega} \right)^{1/2} \times \\ \times \left(\frac{1}{\gamma} \|\mathbf{w}\|^2_{2,\Omega} + \frac{1}{\gamma_*} \|\mathbf{w}\|^2_{2,\partial_2\Omega} \right)^{1/2}.$$

Since (Friederichs type inequality)

$$\|\mathbf{w}\|^2_{2,\Omega} \leq \mathbf{C}^2_{\mathbf{F}}(\Omega)\|\nabla \mathbf{w}\|^2_{2,\Omega}, \qquad \forall \mathbf{w} \in \mathbf{V_0},$$

and (trace inequality)

$$\|\mathbf{w}\|^2_{2,\partial_2\Omega} \leq \mathbf{C}^2_{\mathbf{tr}}(\Omega, \partial_2\Omega)\|\mathbf{w}\|^2_{1,2,\Omega}, \qquad \forall \mathbf{w} \in \mathbf{V_0},$$

we find that

$$\frac{\mathbf{1}}{\gamma}\|\mathbf{w}\|^2_{2,\Omega} + \frac{\mathbf{1}}{\gamma_*}\|\mathbf{w}\|^2_{2,\partial_2\Omega} \leq$$

$$\leq \mathbf{C}^2_{\mathbf{F}}\frac{\mathbf{1}}{\gamma}\|\nabla \mathbf{w}\|^2 + \mathbf{C}^2_{\mathbf{tr}}\frac{\mathbf{1}}{\gamma_*}\left(\|\mathbf{w}\|^2_{2,\Omega} + \|\nabla \mathbf{w}\|^2_{2,\Omega}\right) \leq$$

$$\leq \left(\mathbf{C}^2_{\mathbf{F}}\frac{\mathbf{1}}{\gamma} + \mathbf{C}^2_{\mathbf{tr}}\frac{\mathbf{1}}{\gamma_*}\left(\mathbf{1} + \mathbf{C}^2_{\mathbf{F}}\right)\right)\|\nabla \mathbf{w}\|^2_{2,\Omega}.$$

Therefore, there exist a positive constant $\mathbf{C}_\gamma$ such that

$$\frac{1}{\mathbf{C}_\gamma^2} = \inf_{\mathbf{w}\in\mathbf{V_0}} \frac{\int\limits_\Omega \mathbf{A}\nabla\mathbf{w}\cdot\nabla\mathbf{w}\,\mathbf{dx}}{\frac{1}{\gamma}\|\mathbf{w}\|^2_{\mathbf{2},\Omega} + \frac{1}{\gamma_*}\|\mathbf{w}\|^2_{\mathbf{2},\partial_2\Omega}}.$$

The value of this constant can be estimated numerically by minimizing the above quotient on a sufficiently representative finite dimensional subspace. Besides, if $\mathbf{C_F}$ and $\mathbf{C_{tr}}$ are estimated, then

$$\mathbf{C}_\gamma^2 \le \widehat{\mathbf{C}}_\gamma^2 := \left(\mathbf{C}_\mathbf{F}^2\frac{\mathbf{1}}{\gamma} + \mathbf{C}_\mathbf{tr}^2(\mathbf{1}+\mathbf{C}_\mathbf{F}^2)\frac{\mathbf{1}}{\gamma_*}\right)\mathbf{c}_\mathbf{1}^{-1},$$

so that an upper bound of$\mathbf{C}_\gamma$ is directly computed. Now,

$$|\langle \ell + \mathbf{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle| \le$$
$$\le \widehat{\mathbf{C}}_\gamma \left(\gamma\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|^2_{\mathbf{2},\Omega} + \gamma_*\|\mathbf{y}\cdot\boldsymbol{\nu} + \mathbf{F}\|^2_{\mathbf{2},\partial_2\Omega}\right)^{\mathbf{1/2}} |\!|\!| \nabla\mathbf{w} |\!|\!| .$$

From this estimate, we obtain

$$\| \ell + \Lambda^* \mathbf{y} \|^2 \leq \widehat{C}_\gamma^2 \left(\gamma \| \operatorname{div} \mathbf{y} - \mathbf{f} \|_{2,\Omega}^2 + \frac{\gamma}{\gamma - 1} \| \mathbf{y} \cdot \nu + \mathbf{F} \|_{2,\partial_2\Omega}^2 \right).$$

Consider first the case, in which we simply set $\gamma = \gamma^* = 2$. Then

$$\widehat{C}_{(\gamma=2)}^2 := \widehat{C}_2^2 = \frac{1}{2} \left(C_F^2 + C_{tr}^2 (1 + C_F^2) \right) c_1^{-1},$$

$$\| \ell + \Lambda^* \mathbf{y} \|^2 \leq 2 \widehat{C}_2^2 \left(\| \operatorname{div} \mathbf{y} - \mathbf{f} \|_{2,\Omega}^2 + \| \mathbf{y} \cdot \nu + \mathbf{F} \|_{2,\partial_2\Omega}^2 \right).$$

and we find that

$$\mathbf{M}_\oplus(\mathbf{v}, \beta, \mathbf{y}) = \frac{1+\beta}{2} \int_\Omega (\nabla \mathbf{v} - \mathbf{A}^{-1} \mathbf{y}) \cdot (\mathbf{A} \nabla \mathbf{v} - \mathbf{y}) \, d\mathbf{x} +$$

$$+ \frac{1+\beta}{2\beta} \widehat{C}_2^2 \left(\| \operatorname{div} \mathbf{y} - \mathbf{f} \|_{2,\Omega}^2 + \| \mathbf{y} \cdot \nu + \mathbf{F} \|_{2,\partial_2\Omega}^2 \right). \quad (98)$$

This Majorant gives an upper bound of the deviation for any $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$, $\mathbf{y} \in \mathbf{Q}^*$, and $\beta > 0$.

$\mathsf{M}_\oplus$ for mixed boundary conditions

A more exact estimate is obtained if we define γ by minimizing of the Majorant. Then, we obtain

$$\begin{aligned} \mathbf{]}\, \ell + \Lambda^* \mathsf{y} \,\mathbf{[}^2 \quad &\le \Big(\mathsf{C_F} \|\mathbf{div}\, \mathsf{y} - \mathsf{f}\|_{2,\Omega} + \\ &+ \mathsf{C_{tr}} (1 + \mathsf{C_F^2})^{1/2} \|\mathsf{y}\cdot\nu + \mathsf{F}\|_{2,\partial_2\Omega} \Big)^2 \mathsf{c}_1^{-2} \end{aligned}$$

and

$$\mathsf{M}_\oplus(\mathsf{v}, \beta, \mathsf{y}) = \frac{1+\beta}{2} \int_\Omega (\nabla \mathsf{v} - \mathsf{A}^{-1}\mathsf{y}) \cdot (\mathsf{A}\nabla \mathsf{v} - \mathsf{y})\, \mathbf{dx} +$$

$$+ \frac{1+\beta}{2\beta} \Big(\mathsf{C_F} \|\mathbf{div}\, \mathsf{y} - \mathsf{f}\|_{2,\Omega} + \mathsf{C_{tr}} (1 + \mathsf{C_F^2})^{1/2} \|\mathsf{y}\cdot\nu + \mathsf{F}\|_{2,\partial_2\Omega} \Big)^2 \mathsf{c}_1^{-2}. \quad (99)$$

Majorant vanishes if and only if $\mathsf{v} = \mathbf{u}$ and $\mathsf{y} = \mathsf{A}\nabla\mathbf{u}$, it is continuous with respect to the convergence of v in V and y in Q.

Lower estimates

Lower estimates for the problems considered follow from the general ones obtained in the previous lecture. They have the form

$$\frac{1}{2}\int_\Omega \mathbf{A}\nabla(\mathbf{v}-\mathbf{u})\cdot\nabla(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \geq \mathbf{M}_\ominus(\mathbf{v},\mathbf{w}), \qquad \forall \mathbf{w}\in\mathbf{V_0},$$

where

$$\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}) = -\frac{\mathbf{1}}{\mathbf{2}}\int_\Omega \mathbf{A}\nabla(\mathbf{w}-\mathbf{v})\cdot\nabla\mathbf{w}\,\mathbf{dx} - \int_\Omega \mathbf{f}\mathbf{w}\,\mathbf{dx} - \int_{\partial_2\Omega}\mathbf{F}\mathbf{w}\,\mathbf{ds}.$$

Here $\mathbf{V_0}$ depends on the type of boundary conditions, and the integral over $\partial_2\Omega$ must be eliminated in the case of Dirichlét problem.

Linear elasticity

Classical statement. The classical formulation is as follows:

Find a tensor-valued function $\boldsymbol{\sigma}^*$ (stress) and a vector-valued function $\mathbf{u}$ (displacement) that satisfy the system of equations

$$\begin{aligned}
&\boldsymbol{\sigma}^* = \mathbb{L}\boldsymbol{\varepsilon}(\mathbf{u}) \quad \text{in} \quad \Omega, \quad \text{(Hooke's law)}\\
&\mathbf{div}\boldsymbol{\sigma}^* = \mathbf{f} \text{ in} \quad \Omega, \quad \text{(Equilibrium equation)}\\
&\mathbf{u} = \mathbf{u_0} \quad \text{on} \quad \partial_1\Omega,\\
&\boldsymbol{\sigma}^*\boldsymbol{\nu} + \mathbf{F} = \mathbf{0} \ \text{on} \quad \partial_2\Omega.
\end{aligned}$$

where $\boldsymbol{\varepsilon}(\mathbf{u})$ is a symmetric part of the tensor $\nabla\mathbf{u}$.

Here Ω is a bounded domain with Lipschitz boundary $\partial\Omega$ that consists of two disjoint parts $\partial_1\Omega$ and $\partial_2\Omega$, $|\partial_1\Omega| > 0$, $\mathbf{f}$ and $\mathbf{F}$ are given forces and $\mathbb{L} = \{L_{ijkm}\}$ is the tensor of elasticity constants, which is subject to the conditions

$$\mathbf{C_1}|\varepsilon|^2 \leq \mathbb{L}\varepsilon : \varepsilon \leq \mathbf{C_2}|\varepsilon|^2, \qquad \forall \varepsilon \in \mathbb{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}},$$

and

$$L_{ijkm} = L_{jikm} = L_{kmij}, \qquad L_{ijkm} \in \mathbf{L}^{\infty}(\mathbf{\Omega}).$$

Generalized solution

Let

$$\mathbf{f} \in \mathbf{L}^2(\mathbf{\Omega}, \mathbb{R}^n), \quad \mathbf{F} \in \mathbf{L}^2(\partial_2\mathbf{\Omega}, \mathbb{R}^n).$$

Then, a generalized solution $\mathbf{u} \in \mathbf{V_0} + \mathbf{u_0}$ is defined by the identity

$$\int_{\mathbf{\Omega}} \mathbb{L}\,\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{w})\,\mathbf{dx} + \langle \ell, \mathbf{w} \rangle = \mathbf{0}, \qquad \forall \mathbf{w} \in \mathbf{V_0}, \tag{100}$$

where

$$\langle \ell, \mathbf{w} \rangle = \int_{\mathbf{\Omega}} \mathbf{f} \cdot \mathbf{w}\,\mathbf{dx} + \int_{\partial_2\mathbf{\Omega}} \mathbf{F} \cdot \mathbf{w}\,\mathbf{ds}.$$

Assume that $\mathbf{u}$ is a smooth function and it satisfies the identity

$$\int_{\Omega} \mathbb{L}\,\boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{w})\,\mathbf{dx} + \langle \ell, \mathbf{w} \rangle \;=\; \mathbf{0}, \qquad \forall \mathbf{w} \in \mathbf{V_0},$$

Then,

$$\int_{\Omega} (\mathbf{f} - \mathbf{div}(\mathbb{L}\,\boldsymbol{\varepsilon}(\mathbf{u})) \cdot \mathbf{w}\,\mathbf{dx} + \int_{\partial_2\Omega} \Big((\mathbb{L}\,\boldsymbol{\varepsilon}(\mathbf{u}))\boldsymbol{\nu} + \mathbf{F} \Big) \cdot \mathbf{w}\,\mathbf{ds} \;=\; \mathbf{0},$$
$$\forall \mathbf{w} \in \mathbf{V_0},$$

and we observe that in such a case the equilibrium equation and the Neumann boundary condition are satisfied in the classical sense.

Variational formulation

Note that the relation (100) is the Euler's equation for the functional

$$\mathbf{J(v)} = \frac{\mathbf{1}}{\mathbf{2}} \int_{\mathbf{\Omega}} \mathbb{L}\boldsymbol{\varepsilon}(\mathbf{v}) : \boldsymbol{\varepsilon}(\mathbf{v})\, \mathbf{dx} + < \ell, \mathbf{v} > .$$

Therefore, the respective boundary–value problem may be considered as a minimization problem for $\mathbf{J(v)}$ on the set

$$\mathbf{V_0} := \{\mathbf{v} \in \mathbf{H^1}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{n}}) \mid \mathbf{v} = \mathbf{u_0} \textbf{ on } \partial_1\mathbf{\Omega}\}.$$

To prove existence of a minimizer we must show the coercivity of $\mathbf{J(v)}$ on $\mathbf{V_0}$. The key role in this belongs to the so–called Korn's inequality.

In the Dirichlet problem

$$\mathbf{J}(\mathbf{v}) = \frac{\mathbf{1}}{\mathbf{2}} \int_{\Omega} \mathbb{L}\boldsymbol{\varepsilon}(\mathbf{v}) : \boldsymbol{\varepsilon}(\mathbf{v})\, \mathbf{dx} + < \boldsymbol{\ell}, \mathbf{v} > \geq$$

$$\geq \frac{\mathbf{C_1}}{2} \|\boldsymbol{\varepsilon}(\mathbf{v})\|^2 - \|\mathbf{f}\|\|\mathbf{v}\| =$$

$$= \frac{\mathbf{C_1}}{2} \|\boldsymbol{\varepsilon}(\mathbf{u_0} + \mathbf{w})\|^2 - \|\mathbf{f}\|\|\mathbf{u_0} + \mathbf{w}\| \geq$$

$$\geq \frac{\mathbf{C_1}}{2} (\|\boldsymbol{\varepsilon}(\mathbf{u_0})\| - \|\boldsymbol{\varepsilon}(\mathbf{w})\|)^2 - \|\mathbf{f}\|\|\mathbf{u_0}\| - \|\mathbf{f}\|\|\mathbf{w}\|.$$

Thus, if we can prove that

$$\|\boldsymbol{\varepsilon}(\mathbf{w})\| \geq \mathbf{c}\|\nabla \mathbf{w}\| \quad \forall \mathbf{w} \in \mathring{\mathbf{H}}^{\mathbf{1}}(\boldsymbol{\Omega}),$$

then we would establish the coercivity of $\mathbf{J}$.

Korns's inequality

This inequality is required in various aspects of the mathematical analysis of elasticity problems. In the general form it states the equivalence of two norms:

$$\|\mathbf{w}\|_{1,2,\Omega} := \left(\int_{\Omega} (|\nabla \mathbf{w}|^2 + |\mathbf{w}|^2)\, \mathbf{dx} \right)^{1/2},$$

and

$$\| \mathbf{w} \|_{1,2,\Omega} := \left(\int_{\Omega} (|\varepsilon(\mathbf{w})|^2 + |\mathbf{w}|^2)\, \mathbf{dx} \right)^{1/2}.$$

Korns's inequality in $\overset{\circ}{H}{}^1$

For the functions in $\overset{\circ}{\mathbf{H}}{}^1(\Omega)$ this fact is not difficult to prove. Indeed, for smooth functions

$$\int_\Omega |\varepsilon(\mathbf{w})|^2 d\mathbf{x} = \frac{1}{2}\|\nabla \mathbf{w}\|^2 + \frac{1}{2}\int_\Omega \sum_{ij} w_{i,j} w_{j,i} d\mathbf{x} =$$

$$= \frac{1}{2}\|\nabla \mathbf{w}\|^2 - \frac{1}{2}\int_\Omega \sum_{ij} w_i w_{j,ij} d\mathbf{x} = \frac{1}{2}\|\nabla \mathbf{w}\|^2 + \frac{1}{2}\int_\Omega \sum_{ij} w_{i,i} w_{j,j} d\mathbf{x} =$$

$$= \frac{1}{2}\|\nabla \mathbf{w}\|^2 + \frac{1}{2}\int_\Omega \sum_{i} |w_{i,i}|^2 d\mathbf{x}.$$

Since smooth functions are dense in $\overset{\circ}{\mathbf{H}}{}^1$, we find that

$$\|\nabla \mathbf{w}\| \le \sqrt{2}\|\varepsilon(\mathbf{w})\| \qquad \forall \mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^1(\Omega). \tag{101}$$

By (101) we prove that the energy functional of the elasticity problem for the case of Dirichl'et boundary conditions is coercive, i.e.,

$$\mathbf{J(v_k)} \longrightarrow +\infty, \quad \textbf{as } \|\nabla \mathbf{v_k}\| \longrightarrow +\infty.$$

Rigid deflections

In the analysis of elasticity problems one more notion is often required. It is the so–called *Space of Rigid Deflections* that we denote $\mathbf{RD}(\Omega)$. This space is the kernel of the operator $\varepsilon(\mathbf{w})$, i.e. it contains vector–valued functions $\mathbf{w}$ such that

$$\varepsilon(\mathbf{w}) = \mathbf{0}.$$

It can be defined as follows:

$$\mathbf{RD}(\Omega) := \{\mathbf{w} = \mathbf{w_0} + \omega_0 \mathbf{x} \mid \mathbf{w_0} \in \mathbb{R}^{\mathbf{n}},\ \omega_0 \in \mathbb{M}^{\mathbf{n}\times\mathbf{n}}\},$$

where $\omega_0(\mathbf{w}) = \frac{1}{2}(\nabla \mathbf{w} - (\nabla \mathbf{w})^{\mathbf{T}})$ is a sqew-symmetric tensor associated with "rigid rotations".

Implications of the Korn's inequality

Theorem

Let Ω be a Lipschitz domain and $\partial_1\Omega$ is a nonempty connected part of the boundary. Then,

$$\|\mathbf{u}\|_{1,\mathbf{p},\Omega} \leq \mathbf{C}\left(\int_{\Omega} |\varepsilon(\mathbf{u})|^{\mathbf{p}}\,\mathbf{dx}\right)^{\frac{1}{\mathbf{p}}} \qquad \forall \mathbf{u} \in \mathbf{V_0}, \quad \mathbf{p} \in (\mathbf{1}, \mathbf{2}] \tag{102}$$

Proof. Assume the opposite. Then, for any $m \in \mathbb{N}$ we can find $\mathbf{v}^{(\mathbf{m})}$ such that $\mathbf{v}^{(\mathbf{m})} \in \mathbf{V_0}$ and

$$\|\mathbf{v}^{(\mathbf{m})}\|_{1,\mathbf{p},\Omega} > \mathbf{m}\left(\int_{\Omega} |\varepsilon(\mathbf{v}^{(\mathbf{m})})|^{\mathbf{p}}\,\mathbf{dx}\right)^{\frac{1}{\mathbf{p}}}.$$

Set $\mathbf{w^{(m)}} = \frac{\mathbf{v^{(m)}}}{\|\mathbf{v^{(m)}}\|_{\mathbf{1,p,\Omega}}}$, then

$$\|\mathbf{w^{(m)}}\|_{\mathbf{1,p,\Omega}} = \mathbf{1} \qquad \text{and} \qquad \frac{\mathbf{1}}{\mathbf{m}} \geq \left(\int_{\Omega} |\boldsymbol{\varepsilon}(\mathbf{w^{(m)}}|^{\mathbf{p}} \, \mathbf{dx} \right)^{\frac{\mathbf{1}}{\mathbf{p}}}.$$

Therefore,

$$\begin{aligned} \mathbf{w^{(m)}} &\rightharpoonup \mathbf{w} \quad \text{in} \quad \mathbf{W^1_p}(\Omega, \mathbf{R^n}), \\ \mathbf{w^{(m)}} &\to \mathbf{w} \quad \in \quad \mathbf{L^p}(\Omega, \mathbf{R^n}), \\ \|\boldsymbol{\varepsilon}(\mathbf{w^{(m)}})\|_{\mathbf{p},\Omega} &\to \mathbf{0} \quad \text{in} \quad \mathbf{L^p}(\Omega, \mathbf{R^n}). \end{aligned}$$

From here we conclude that $\boldsymbol{\varepsilon}(\mathbf{w}) = \mathbf{0}$.

Indeed, by the fact that a norm is weakly lower semicontinuous, we have

$$\mathbf{0} = \liminf_{\mathbf{m}} \|\varepsilon(\mathbf{w}^{(\mathbf{m})})\|_{\mathbf{p},\Omega} \geq \|\varepsilon(\mathbf{w})\|_{\mathbf{p},\Omega}.$$

Thus, $\mathbf{w} \in \mathbf{RD}(\Omega) \cap \mathbf{V}$. There is only one such a function: $\mathbf{w} = \mathbf{0}$. It means that $\mathbf{w}^{(\mathbf{m})} \longrightarrow \mathbf{0}$ in $\mathbf{L}^{\mathbf{p}}$. Now, we apply Korn's inequality

$$\|\mathbf{w}(\mathbf{m})\|_{\mathbf{1},\mathbf{p},\Omega} \leq \mathbf{C} \left(\int_{\Omega} \left(|\varepsilon(\mathbf{w}^{(\mathbf{m})})|^{\mathbf{p}} + |\mathbf{w}^{(\mathbf{m})}|^{\mathbf{p}} \right) \mathbf{dx} \right)^{\frac{\mathbf{1}}{\mathbf{p}}} \underset{\mathbf{m}\to\infty}{\longrightarrow} \mathbf{0},$$

which shows that $\|\mathbf{w}(\mathbf{m})\|_{\mathbf{1},\mathbf{p},\Omega}$ tends to zero. But for any $\mathbf{m}$ $\|\mathbf{w}^{(\mathbf{m})}\|_{\mathbf{p},\mathbf{1},\Omega} = \mathbf{1}$, so that such a behavior is impossible. We have arrived at a contradiction that proves the Theorem.

Another similar result is required for the Neumann problem. Define the set

$$\mathbf{V} = \left\{ \mathbf{v} \in \mathbf{W}^1_{\mathbf{p}}(\Omega) \mid \int_\Omega \mathbf{v} \cdot \mathbf{w} \, \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{RD}(\Omega) \right\} .$$

Theorem

Let Ω be a bounded domain with Lipschitz boundary $\partial\Omega$. Then

$$\|\mathbf{u}\|_{\mathbf{1},\mathbf{p},\Omega} \leq \mathbf{C} \left(\int_\Omega |\varepsilon(\mathbf{u})|^{\mathbf{p}} \, \mathbf{dx} \right)^{\frac{\mathbf{1}}{\mathbf{p}}} \quad \forall \mathbf{u} \in \mathbf{V}. \tag{103}$$

Proof. By the same arguments as before, we obtain a sequence $\mathbf{w}^{(m)} \in \mathbf{V}$ such that

$$\mathbf{w}^{(m)} \rightharpoonup \mathbf{w} \quad \text{in} \quad \mathbf{W}^1_p(\Omega, \mathbf{R}^n),$$
$$\mathbf{w}^{(m)} \to \mathbf{w} \quad \in \quad \mathbf{L}^p(\Omega, \mathbf{R}^n),$$
$$\|\boldsymbol{\varepsilon}(\mathbf{w}^{(m)})\|_{p,\Omega} \to \mathbf{0} \quad \text{in} \quad \mathbf{L}^p(\Omega, \mathbf{R}^n).$$

By the arguments similar to those in the previous Theorem, we find that $\boldsymbol{\varepsilon}(\mathbf{w}) = \mathbf{0}$ and, thus, $\mathbf{w} \in \mathbf{RD}(\Omega)$. In addition, for any $\bar{\mathbf{w}} \in \mathbf{RD}$, we have

$$\mathbf{0} = \int_\Omega \mathbf{w}^{(m)} \cdot \bar{\mathbf{w}} \, \mathbf{dx} = \int_\Omega \mathbf{w} \cdot \bar{\mathbf{w}} \, \mathbf{dx}.$$

But $\mathbf{w} \in \mathbf{RD}$, so that $\|\mathbf{w}\| = \mathbf{0}$, and by applying Korn's inequality we prove that $\|\mathbf{w}^{(m)}\|_{1,p,\Omega}$ tends to zero, what leads to a contradiction.

Estimates of deviations

Let $\mathbf{v}$ and $\mathbf{y}$ be some approximations of $\mathbf{u}$ and $\boldsymbol{\sigma}^*$. Estimates of $\mathbf{v}-\mathbf{u}$ and $\mathbf{y}-\boldsymbol{\sigma}^*$ follow from the general scheme if we set

$$\mathbf{U} = \mathbf{L}^2(\Omega, \mathbb{M}^{n\times n}_s), \qquad \mathbf{V} = \mathbf{H}^1(\Omega, \mathbb{R}^n),$$
$$\mathbf{V_0} = \{\mathbf{w} \in \mathbf{V} \,|\, \mathbf{w} = \mathbf{0} \text{ on } \partial_1\Omega\},$$
$$\| \mathbf{y} \|^2 = \int_\Omega \mathbb{L}\mathbf{y} : \mathbf{y}\, \mathbf{dx}, \qquad \| \mathbf{y} \|^2_* = \int_\Omega \mathbb{L}^{-1}\mathbf{y} : \mathbf{y}\, \mathbf{dx},$$

and $\boldsymbol{\Lambda}\mathbf{v} = \boldsymbol{\varepsilon}(\mathbf{v}) := \frac{1}{2}\left(\nabla\mathbf{v} + (\nabla\mathbf{v})^{\mathsf{T}}\right)$. In this case,

$$\langle \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle = \int_\Omega \mathbf{y} : \boldsymbol{\varepsilon}(\mathbf{w})\, \mathbf{dx}, \qquad \forall \mathbf{w} \in \mathbf{V_0},$$

Now $\mathbf{y}$ is a tensor-valued function and $\mathbf{y}\boldsymbol{\nu} = \mathbf{y_{ij}}\boldsymbol{\nu_j}$ is a vector–function defined on $\partial\Omega$.
If

$$\mathbf{y} \in \mathbf{Q}^* := \{\mathbf{y} \in \mathbf{Y}^* \mid \mathbf{div y} \in \mathbf{L}^2(\boldsymbol{\Omega}, \mathbb{M}^{n\times n}),\ \mathbf{y}\boldsymbol{\nu} \in \mathbf{L}^2(\partial_2\boldsymbol{\Omega}, \mathbb{R}^n)\}.$$

then

$$\langle \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle = -\int_{\boldsymbol{\Omega}} \mathbf{div}\, \mathbf{y} \cdot \mathbf{w}\, \mathbf{dx} + \int_{\partial_2\boldsymbol{\Omega}} (\mathbf{y}\boldsymbol{\nu}) \cdot \mathbf{w}\, \mathbf{d\Gamma}$$

so that

$$\boldsymbol{\Lambda}^*\mathbf{y} = \{-\mathbf{div y}\ |_{\boldsymbol{\Omega}},\quad (\mathbf{y}\boldsymbol{\nu})\ |_{\partial_2\boldsymbol{\Omega}}\}.$$

Upper estimates

By applying the general estimate, we obtain the following upper estimate:

$$\frac{1}{2}\int_\Omega \mathbb{L}\,\varepsilon(\mathbf{v}-\mathbf{u}) : \varepsilon(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \le \mathbf{M}_\oplus(\mathbf{v},\beta,\mathbf{y}),$$

where

$$\mathbf{M}_\oplus(\mathbf{v},\beta,\mathbf{y}) = \frac{1+\beta}{2}\mathbf{D}(\varepsilon\mathbf{v},\mathbf{y}) + \frac{1+\beta}{2\beta}\,\mathbf{[}\,\mathbf{\Lambda}^*\mathbf{y}+\ell\,\mathbf{]}^2$$

and

$$\mathbf{D}(\varepsilon(\mathbf{v}),\mathbf{y}) = \frac{1}{2}\int_\Omega \Big(\mathbb{L}\varepsilon(\mathbf{v}) : \varepsilon(\mathbf{v})\mathbb{L}^{-1}\mathbf{y} : \mathbf{y} - 2\varepsilon(\mathbf{v}) : \mathbf{y}\Big)\,\mathbf{dx} =$$

$$= \int_\Omega (\varepsilon(\mathbf{u})-\mathbb{L}^{-1}\mathbf{y}) : (\mathbb{L}\,\varepsilon(\mathbf{u})-\mathbf{y})\,\mathbf{dx}.$$

If $\mathbf{y} \in \mathbf{Q}^*$, then

$$\| \Lambda^* \mathbf{y} + \ell \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\langle \Lambda^* \mathbf{y} + \ell, \mathbf{w} \rangle}{\| \Lambda \mathbf{w} \|} =$$

$$= \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{y} : \varepsilon(\mathbf{w}) + \mathbf{f} \cdot \mathbf{w}) \, \mathbf{dx} + \int_{\partial_2 \Omega} \mathbf{F} \cdot \mathbf{w} \, \mathbf{ds}}{\| \varepsilon(\mathbf{w}) \|} =$$

$$= \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{f} - \mathbf{div}\, \mathbf{y}) \cdot \mathbf{w} \, \mathbf{dx} + \int_{\partial_2 \Omega} (\mathbf{F} + \mathbf{y}\nu) \cdot \mathbf{w} \, \mathbf{ds}}{\| \varepsilon(\mathbf{w}) \|} \leq$$

$$\leq \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\|\mathbf{f} - \mathbf{div}\, \mathbf{y}\|_{2,\Omega} \|\mathbf{w}\|_{2,\Omega} + \|\mathbf{F} + \mathbf{y}\nu\|_{\partial_2 \Omega} \|\mathbf{w}\|_{\partial_2 \Omega}}{\| \varepsilon(\mathbf{w}) \|}.$$

Let $\mathbf{C_\Omega}$ be a constant in the inequality

$$\int_\Omega |\mathbf{w}|^2 \, \mathbf{dx} + \int_{\partial_2\Omega} |\mathbf{w}|^2 \, \mathbf{ds} \le \mathbf{C}_\Omega^2 \|\boldsymbol{\varepsilon}(\mathbf{w})\|_\Omega^2, \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

Note that the existence of such a constant follows from the Korn's inequality. Indeed, the inequality

$$\int_\Omega |\mathbf{w}|^2 \, \mathbf{dx} + \int_{\partial_2\Omega} |\mathbf{w}|^2 \, \mathbf{ds} \le \hat{\mathbf{C}}_\Omega^2 \|\nabla(\mathbf{w})\|_\Omega^2, \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

for the tensor–gradient $\nabla(\mathbf{w})$ follows from the Friederichs type inquality for the vector–valued functions and the respective trace theorems. By (102) we recall that for the functions in $\mathbf{V_0}$

$$\|\nabla(\mathbf{w})\|_\Omega \le \mathbf{C} \|\boldsymbol{\varepsilon}(\mathbf{w})\|_\Omega$$

with a certain constant $\mathbf{C}$ and the estimate follows.

In practice, values of $\mathbf{C_\Omega}$ can be estimated by minimizing the quotient

$$\frac{\|\varepsilon(\mathbf{w})\|^2_\Omega}{\int\limits_\Omega |\mathbf{w}|^2\,\mathbf{dx} + \int\limits_{\partial_2\Omega} |\mathbf{w}|^2\,\mathbf{ds}}$$

over sufficiently representative finite dimensional space $\mathbf{V_{0h}} \subset \mathbf{V_0}$.

Let us now return to finding an upper bound of the quantity $\mathbf{[}\,\mathbf{\Lambda}^*\mathbf{y} + \ell\,\mathbf{]}$. By the inequality $ab + cd \le \sqrt{a^2+c^2}\sqrt{b^2+d^2}$, we obtain

$$\mathbf{[}\,\mathbf{\Lambda}^*\mathbf{y} + \ell\,\mathbf{]} \le$$

$$\le \left(\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|^2_\Omega + \|\mathbf{F} + \mathbf{y}\nu\|^2_{\partial_2\Omega}\right)^{1/2} \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\left(\|\mathbf{w}\|^2_\Omega + \|\mathbf{w}\|^2_{\partial_2\Omega}\right)^{1/2}}{|||\ \varepsilon(\mathbf{w})\ |||} \le$$

$$\le \mathbf{C_\Omega c_1^{-1/2}} \left(\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|^2_\Omega + \|\mathbf{F} + \mathbf{y}\nu\|^2_{\partial_2\Omega}\right)^{1/2}$$

Error Majorant for mixed boundary conditions

Hence, we arrive at the Majorant $\mathbf{M}_{\oplus}$:

$$\mathbf{M}_{\oplus}(\varepsilon(\mathbf{v}),\mathbf{y}) = \frac{1+\beta}{2}\int_{\Omega}(\varepsilon(\mathbf{v})-\mathbb{L}^{-1}\mathbf{y}):(\mathbb{L}\,\varepsilon(\mathbf{v})-\mathbf{y})\,\mathbf{dx}+$$

$$+\frac{1+\beta}{2\beta c_1}\,\mathbf{C}_{\Omega}^2\left(\|\mathbf{div}\,\mathbf{y}-\mathbf{f}\|_{\Omega}^2+\|\mathbf{F}+\mathbf{y}\nu\|_{\partial_2\Omega}^2\right). \quad (104)$$

It has a clear physical meaning. The first term of $\mathbf{M}_{\oplus}$ is nonnegative and vanishes if and only if

$$\mathbf{y} = \mathbb{L}\varepsilon(\mathbf{v}).$$

It penalizes violations of the **Hooke's law**. The meaning of the second term is obvious: it contains $\mathbf{L}^2$-norms of other two relations, which gives errors in the **equilibrium equation** and **boundary condition** for the stress tensor.

Thus, the majorant not only gives an idea of the overall value of the error, but also shows its physically sensible parts.

Let $\{\mathbf{Y}_k^*\} \subset \mathbf{H}^1(\Omega, \mathbb{M}^{n\times n})$ be a collection of finite-dimensional subspaces that satisfy the limit density condition. Then, (104) generates a sequence of computable upper bounds

$$\mathbf{M}_{\oplus}^{k} = \inf_{\substack{\mathbf{y}\in\mathbf{Y}_k^* \\ \beta\in\mathbb{R}_+}} \Big\{ \frac{1+\beta}{2} \int_\Omega \big(\mathbb{L}\,\varepsilon(\mathbf{v}) : \varepsilon(\mathbf{v}) + \mathbb{L}^{-1}\mathbf{y} : \mathbf{y} - 2\varepsilon(\mathbf{v}) : \mathbf{y} \big)\, dx + \frac{1+\beta}{2\beta c_1} \mathbf{C}_\Omega^2 \big(\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|_\Omega^2 + \|\mathbf{F} + \mathbf{y}\nu\|_{\partial_2\Omega}^2 \big) \Big\},$$

which tends to the exact value of the error.

Lower estimates

Lower estimates also follow from the general theory. We have

$$\frac{1}{2}\int_\Omega \mathbb{L}\,\varepsilon(\mathbf{v}-\mathbf{u}) : \varepsilon(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \geq \mathbf{M}_\ominus(\mathbf{v},\mathbf{w}), \qquad \forall \mathbf{w} \in \mathbf{V_0},$$

where

$$\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}) = -\frac{1}{2}\int_\Omega \mathbb{L}\,\varepsilon(\mathbf{w}) : \varepsilon(\mathbf{w})\,\mathbf{dx} - \int_\Omega \mathbb{L}\,\varepsilon(\mathbf{v}) : \varepsilon(\mathbf{w})\,\mathbf{dx} - \int_\Omega \mathbf{f}\cdot\mathbf{w}\,\mathbf{dx} - \int_{\partial_2\Omega} \mathbf{F}\cdot\mathbf{w}\,\mathbf{ds}.$$

By the same arguments as for the diffusion equation one can prove that

$$\frac{1}{2}\int_\Omega \mathbb{L}\,\varepsilon(\mathbf{v}-\mathbf{u}) : \varepsilon(\mathbf{v}-\mathbf{u})\,\mathbf{dx} = \sup_{\mathbf{w}\in\mathbf{V_0}} \mathbf{M}_\ominus(\mathbf{v},\mathbf{w_k}).$$

By the maximization the functional $\mathbf{M}_\ominus$ on a sequence of finite-dimensional spaces $\mathbf{V_{0k}} \subset \mathbf{V_0}$, we obtain a sequence of computable lower bounds

$$\mathbf{M}_\ominus^{\mathbf{k}} = \sup_{\mathbf{w} \in \mathbf{V_{0k}}} \mathbf{M}_\ominus(\mathbf{v}, \mathbf{w_k}).$$

If the spaces $\mathbf{V_{0k}}$ satisfy the limit density condition stated, then the sequence of numbers $\{\mathbf{M}_\ominus\}$ tends to $\frac{1}{2}|||\boldsymbol{\varepsilon}(\mathbf{v}-\mathbf{u})|||^2$.

Linear elliptic equations of the fourth order

Now, we consider the problem

$$\nabla \cdot \nabla \cdot (\mathbf{B} \nabla \nabla \mathbf{u}) = \mathbf{f} \quad \text{in} \quad \Omega, \tag{105}$$

$$\mathbf{u} = \frac{\partial \mathbf{u}}{\partial \boldsymbol{\nu}} = \mathbf{0} \quad \text{on} \quad \partial\Omega. \tag{106}$$

Here $\Omega \subset \mathbb{R}^2$, ν denotes the outward unit normal to the boundary, and $\mathbf{B} = \{\mathbf{b_{ijkl}}\} \in \mathcal{L}(\mathbb{M}_{\mathbf{s}}^{\mathbf{2\times 2}}, \mathbb{M}_{\mathbf{s}}^{\mathbf{2\times 2}})$. We assume that $\mathbf{b_{ijkl}} = \mathbf{b_{jikl}} = \mathbf{b_{klij}}$,

$$\alpha_1 |\boldsymbol{\eta}|^2 \leq \mathbf{B}\boldsymbol{\eta} : \boldsymbol{\eta} \leq \alpha_2 |\boldsymbol{\eta}|^2, \quad \forall \boldsymbol{\eta} \in \mathbb{M}_{\mathbf{s}}^{\mathbf{2\times 2}},$$

and

$$\mathbf{f} \in \mathbf{L}^2(\Omega), \quad \mathbf{b_{ijkl}} \in \mathbf{L}^\infty(\Omega).$$

To apply the general scheme, we set

$$\mathbf{U} = \mathbf{L}^2(\Omega, \mathbb{M}_s^{2\times 2}), \qquad \mathbf{V} = \mathbf{H}^2(\Omega),$$
$$\mathbf{V_0} = \{\mathbf{w} \in \mathbf{V} \mid \mathbf{w} = \frac{\partial \mathbf{w}}{\partial \nu} = \mathbf{0} \text{ on } \partial\Omega\},$$

and define $\mathbf{\Lambda}$ as the Hessian operator. Now, the basic integral identity has the form

$$\int_\Omega \mathbf{B}\nabla\nabla\mathbf{u} : \nabla\nabla\mathbf{w}\,\mathbf{dx} = \int_\Omega \mathbf{f}\mathbf{w}\,\mathbf{dx} \quad \forall \mathbf{w} \in \mathbf{V_0}. \tag{107}$$

By $\mathbf{B}^{-1}$ we denote the inverse tensor, which satisfies the double inequality

$$\boldsymbol{\alpha}_2^{-1}|\boldsymbol{\eta}|^2 \le \mathbf{B}^{-1}\boldsymbol{\eta} : \boldsymbol{\eta} \le \boldsymbol{\alpha}_1^{-1}|\boldsymbol{\eta}|^2, \quad \forall \boldsymbol{\eta} \in \mathbb{M}_s^{2\times 2},$$

The spaces $\mathbf{Y}$ and $\mathbf{Y}^*$ are equipped with norms

$$\| \mathbf{y} \|^2 = \int_\Omega \mathbf{B}\mathbf{y} : \mathbf{y} \, \mathbf{dx}; \quad \| \mathbf{y} \|_*^2 = \int_\Omega \mathbf{B}^{-1}\mathbf{y} : \mathbf{y} \, \mathbf{dx},$$

$$\langle \ell, \mathbf{w} \rangle = - \int_\Omega \mathbf{f}\mathbf{w} \, \mathbf{dx},$$

and

$$\mathbf{Q}_\ell^* = \{ \mathbf{y} \in \mathbf{Y}^* \; \Big| \; \int_\Omega \mathbf{y} : \nabla\nabla\mathbf{w} \, \mathbf{dx} = \int_\Omega \mathbf{f}\mathbf{w} \, \mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0} \}.$$

Since

$$\|\nabla\nabla\mathbf{w}\| \geq \boldsymbol{\alpha_3} \|\mathbf{w}\|_{2,2,\Omega} \quad \forall \mathbf{w} \in \mathbf{V_0},$$

we have the required version of the coercivity condition

$$\|\boldsymbol{\Lambda}\mathbf{w}\| \geq \mathbf{c_3} \|\mathbf{w}\|_{\mathbf{V}}.$$

Problem (105) and (106) is associated with two variational problems.

Problem $\mathcal{P}$. Find $\mathbf{u} \in \mathbf{V_0}$ such that

$$\mathbf{J}(\mathbf{u}) = \inf_{\mathbf{v} \in \mathbf{V_0}} \mathbf{J}(\mathbf{v}),$$

where

$$\mathbf{J}(\mathbf{v}) = \frac{1}{2} \int_\Omega \mathbf{B} \nabla\nabla \mathbf{v} : \nabla\nabla \mathbf{v} \, \mathbf{dx} - \int_\Omega \mathbf{f} \mathbf{w} \, \mathbf{dx}.$$

Problem $\mathcal{P}^*$. Find $\mathbf{p} \in \mathbf{Q}_\ell^*$ such that

$$\mathbf{I}^*(\mathbf{p}) = \sup_{\forall \mathbf{q} \in \mathbf{Q}_\ell^*} \mathbf{I}^*(\mathbf{q}),$$

where

$$\mathbf{I}^*(\mathbf{q}) = -\frac{1}{2} \int_\Omega \mathbf{B}^{-1} \mathbf{q} : \mathbf{q} \, \mathbf{dx}.$$

Two basic relations for the deviations that we derived in the previous Lecture now come in the form:

$$\| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 + \| \mathbf{q}-\mathbf{p} \|_*^2 = 2(J(\mathbf{v}) - I^*(\mathbf{q})), \tag{108}$$

and

$$\| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 + \| \mathbf{q}-\mathbf{p} \|_*^2 = 2D(\nabla\nabla\mathbf{v}, \mathbf{q}) = \\ = \int_\Omega \left(\mathbf{B}\nabla\nabla\mathbf{v} : \nabla\nabla\mathbf{v} + \mathbf{B}^{-1}\mathbf{q} : \mathbf{q} - 2\nabla\nabla\mathbf{v} : \mathbf{q} \right) \mathbf{dx}. \tag{109}$$

They hold for any $\mathbf{v} \in \mathbf{V_0}$ and $\mathbf{q} \in \mathbf{Q}_\ell^*$.
Also, from the general theory it readily follows the first a posteriori estimate:

$$\frac{1}{2} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 \le (1+\beta) D(\nabla\nabla\mathbf{v}, \mathbf{y}) + \left(1 + \frac{1}{\beta}\right) \frac{\mathbf{d}_\ell^2(\mathbf{y})}{2}, \tag{110}$$

where $\mathbf{d}_\ell^2(\mathbf{y}) = \inf_{\mathbf{q} \in \mathbf{Q}_\ell^*} \| \mathbf{q}-\mathbf{y} \|_*^2$.

Note that

$$\int_\Omega \mathbf{y} : \nabla\nabla \mathbf{w}\, \mathbf{dx} = \int_\Omega (\mathbf{divdiv}\, \mathbf{y}) \mathbf{w}\, \mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0},$$

so that $\mathbf{\Lambda}^* : \mathbf{Y}^* \to \mathbf{H}^{-2}(\mathbf{\Omega})$ is the operator **divdiv**.
Next,

$$\langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_\Omega (\mathbf{y} : \nabla\nabla \mathbf{w} - \mathbf{fw})\, \mathbf{dx}$$

and, therefore,

$$\mathbf{d}_\ell^2(\mathbf{y}) = \mathbf{[}\!\mathbf{[}\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,\mathbf{]}\!\mathbf{]} = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{y} : \nabla\nabla \mathbf{w} - \mathbf{fw})\, \mathbf{dx}}{\| \nabla\nabla \mathbf{w} \|}.$$

If

$$\mathbf{y} \in \mathbf{H}(\mathbf{divdiv}, \Omega) := \left\{ \mathbf{y} \in \mathbf{L}^2(\Omega, \mathbb{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}}) \mid \mathbf{divdiv}\, \mathbf{y} \in \mathbf{L}^2(\Omega) \right\},$$

then this quantity is estimated by the relation

$$\mathbf{[}\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,\mathbf{]} \le \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\|\mathbf{divdiv}\, \mathbf{y} - \mathbf{f}\|_\Omega \|\mathbf{w}\|_\Omega}{\| \nabla\nabla \mathbf{w} \|} \le$$

$$\le \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\|\mathbf{divdiv}\, \mathbf{y} - \mathbf{f}\|_\Omega \|\mathbf{w}\|_\Omega}{\alpha_1 \|\nabla\nabla \mathbf{w}\|} \le \frac{\mathbf{C_{1\Omega}}}{\alpha_1} \|\mathbf{divdiv}\, \mathbf{y} - \mathbf{f}\|_\Omega,$$

in which $\mathbf{C_{1\Omega}}$ is a constant in the inequality

$$\|\mathbf{w}\|_\Omega \le \mathbf{C_{1\Omega}} \|\nabla\nabla \mathbf{w}\|_\Omega \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

Now, we obtain the first variant of a posteriori estimate for the biharmonic type problem.

First a posteriori estimate

$$\frac{1}{2} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 \leq (1+\beta)\mathbf{D}(\nabla\nabla\mathbf{v}, \mathbf{y}) + \\ + \left(1+\frac{1}{\beta}\right) \frac{\mathbf{C}_{1\Omega}^2}{2\alpha_1^2} \|\mathbf{divdiv}\,\mathbf{y} - \mathbf{f}\|_\Omega^2, \quad (111)$$

Here, $\mathbf{y}$ is an arbitrary tensor-valued function from $\mathbf{H}(\mathbf{div}\,\mathbf{div}, \Omega)$ and β is a positive real number. However, this is rather demanding in relation to the dual variable $\mathbf{y}$ (which must have square summable **divdiv**). To avoid technical difficulties that rises from this condition, we estimate the negative norm in a different way.

$$\| \ell + \Lambda^* \mathbf{y} \| = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{y} : \nabla\nabla \mathbf{w} - \mathbf{f}\mathbf{w})\, \mathbf{dx}}{\| \nabla\nabla \mathbf{w} \|} =$$

$$= \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{y} : \nabla\nabla \mathbf{w} + \boldsymbol{\eta} \cdot \nabla \mathbf{w} + \mathbf{div}\boldsymbol{\eta}\mathbf{w} - \mathbf{f}\mathbf{w})\, \mathbf{dx}}{\| \nabla\nabla w \|} =$$

$$\frac{\int_\Omega (-\mathbf{div}\mathbf{y} \cdot \nabla \mathbf{w} + \boldsymbol{\eta} \cdot \nabla \mathbf{w} + \mathbf{div}\boldsymbol{\eta}\mathbf{w} - \mathbf{f}\mathbf{w})\, \mathbf{dx}}{\| \nabla\nabla w \|} \leq$$

$$\leq \frac{\mathbf{C_{2\Omega}}}{\alpha_1} \|\mathbf{div}\, \mathbf{y} - \boldsymbol{\eta}\|_\Omega + \frac{\mathbf{C_{1\Omega}}}{\alpha_1} \|\mathbf{div}\, \boldsymbol{\eta} - \mathbf{f}\|_\Omega .$$

Here, $\boldsymbol{\eta}$ is an arbitrary vector-valued function from $\mathbf{H(div, \Omega)}$ and $\mathbf{C_{2\Omega}}$ is a constant in the inequality

$$\|\nabla \mathbf{w}\|_\Omega \leq \mathbf{C_{2\Omega}} \|\nabla\nabla \mathbf{w}\|_\Omega \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

Second a posteriori estimate

Then, we arrive at the estimate

$$\frac{1}{2} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 \leq (1+\beta)\mathbf{D}(\nabla\nabla\mathbf{v}, \mathbf{y}) + \\ + \left(1+\frac{1}{\beta}\right) \frac{1}{2\alpha_1^2} \left(\mathbf{C}_{2\Omega}\|\mathbf{div}\,\mathbf{y}-\eta\|_\Omega + \mathbf{C}_{1\Omega}\|\mathbf{div}\,\eta-\mathbf{f}\|_\Omega\right)^2, \quad (112)$$

in which $\mathbf{y} \in \boldsymbol{\Sigma}_{\mathbf{div}}(\Omega)$ and $\eta \in \mathbf{H}(\mathbf{div}, \Omega)$.

This estimate was obtained in
P. Neittaanmäki and S. Repin. A posteriori error estimates for boundary-value problems related to the biharmonic operator, *East-West J.Numer. Math.*, 9(2001)

Note that

$$\|\mathbf{w}\| \leq \mathbf{C_F}\|\nabla\mathbf{w}\|_\Omega \leq \mathbf{C_F C_{2\Omega}}\|\nabla\nabla\mathbf{w}\|_\Omega \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

where $\mathbf{C_F}$ is a constant in the Friederichs inequality. Therefore, $\mathbf{C_{1\Omega}} \leq \mathbf{C_F C_{2\Omega}}$. In view of this, we obtain a slightly different form of the deviation estimate:

$$\frac{\mathbf{1}}{\mathbf{2}} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 \leq (\mathbf{1}+\beta)\mathbf{D}(\nabla\nabla\mathbf{v}, \mathbf{y}) + \left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\frac{\mathbf{C_{2\Omega}^2}}{\mathbf{2}\alpha_1^2}\left(\|\mathbf{div}\,\mathbf{y}-\eta\|_\Omega + \mathbf{C_F}\|\mathbf{div}\,\eta-\mathbf{f}\|_\Omega\right)^2, \quad (113)$$

For boundary conditions of other types, the deviation majorants can be derived by arguments similar to those used in Lecture 6.

Lower estimates of the deviation from u

Lower estimates follow from the general estimate discussed in Lecture 5. We have

$$\frac{1}{2} \| \nabla\nabla(\mathbf{v} - \mathbf{w}) \|^2 \geq \mathbf{M}_{\ominus}(\mathbf{v}, \mathbf{w}) \quad \mathbf{w} \in \mathbf{V_0}, \tag{114}$$

where

$$\mathbf{M}_{\ominus}(\mathbf{v}, \mathbf{w}) := -\frac{1}{2} \| \nabla\nabla\mathbf{w} \|^2 - \int_{\Omega} (\mathbf{B}\nabla\nabla\mathbf{v} : \nabla\nabla\mathbf{w} - \mathbf{f}\mathbf{w}) d\mathbf{x}.$$

Lecture 5

A POSTERIORI ESTIMATES IN NON-ENERGY QUANTITIES

If we cam estimate the energy error norm, then we can estimate all other quantities subject to it.

General Principle

If

$$\| \mathbf{u} - \mathbf{v} \| \leq \mathbf{M}_{\oplus}$$

and

$$\phi(\mathbf{u} - \mathbf{v}) \leq \mathbf{C} \| \mathbf{u} - \mathbf{u_a} \|$$

then

$$\phi(\mathbf{u} - \mathbf{v}) \leq \mathbf{C}\,\mathbf{M}_{\oplus}$$

Example

By the embedding theorems we know that for any Lipschitz domain

If $q \geq 1$, and $1 + \frac{n}{q} \geq \frac{n}{2}$, then $\mathbf{W^{1,2}(\Omega)}$ is continuously embedded in $\mathbf{L^q(\Omega)}$, i.e.,

$$\|\mathbf{w}\|_{\mathbf{q},\Omega} \leq \mathbf{C(\Omega, q)}\|\mathbf{w}\|_{\mathbf{1,2},\Omega}$$

If $1 > \frac{n}{2}$, then $\mathbf{W^{1,2}(\Omega)}$ is compactly embedded in $\mathbf{C(\overline{\Omega})}$, i.e.,

$$\|\mathbf{w}\|_{\mathbf{C(\Omega)}} \leq \hat{\mathbf{C}}\mathbf{(\Omega)}\|\mathbf{w}\|_{\mathbf{1,2},\Omega}$$

If the conditions of above theorems hold, then we obtain

$$\|\mathbf{u}-\mathbf{v}\|_{\mathbf{q},\Omega} \leq \mathbf{C}(\Omega,\mathbf{q})\mathbf{M}_{\oplus}(\mathbf{v},\mathbf{y})$$

Note that if $\mathbf{n}=\mathbf{2}$ then the above relation holds for any finite $\mathbf{q}$.

$$\mathrm{esssup}_{\mathbf{x}\in\Omega}\|\mathbf{u}(\mathbf{x})-\mathbf{v}(\mathbf{x})\| \leq \mathbf{C}(\Omega,\mathbf{q})\mathbf{M}_{\oplus}(\mathbf{v},\mathbf{y})$$

ERROR ESTIMATES IN TERMS OF GOAL-ORIENTED FUNCTIONALS

$$\ell(\mathbf{u}-\mathbf{v}) \leq \mathbf{M}_{\ell\oplus}$$

Example:

$$< \ell, \mathbf{u}-\mathbf{v} >= \int_{\Omega} \varphi(\mathbf{u}-\mathbf{v})\ \mathbf{dx},$$

where the weight φ is a locally supported function.

Let us consider the methods of goal–oriented error control on the paradigm of our Basic Linear Problem.

$$(\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda w}) = < \mathbf{f}, \mathbf{w} >, \quad \forall \mathbf{w} \in \mathbf{V_0}. \tag{115}$$

$\mathcal{A} \in \mathcal{L}(\mathbf{U}, \mathbf{U})$,
$\Lambda \in \mathcal{L}(\mathbf{V}, \mathbf{U})$,
$\mathbf{V}$ is a Banach space, $\mathbf{V_0}$ is a subspace of $\mathbf{V}$, $f \in \mathbf{V_0^*}$,
$\mathbf{U}$ is a Hilbert space equipped with scalar product $(\cdot, \cdot)$,
$< \cdot, \cdot >$ denotes the duality pairing of $\mathbf{V_0}$ and $\mathbf{V_0^*}$.

We assume that the operators $\mathcal{A}$ and Λ satisfy the conditions

$$\mathbf{c_1}\|\mathbf{y}\|^2_{\mathbf{U}} \leq (\mathcal{A}\mathbf{y}, \mathbf{y}) \leq \mathbf{c_2}\|\mathbf{y}\|^2_{\mathbf{U}}, \forall \mathbf{y} \in \mathbf{U}, \tag{116}$$

$$\mathbf{c_3}\|\mathbf{w}\|_{\mathbf{V}} \leq \|\mathbf{\Lambda w}\|_{\mathbf{U}}, \quad \forall \mathbf{w} \in \mathbf{V_0}, \tag{117}$$

Our aim is to derive estimates of the quantity

$$< \ell, \mathbf{u} - \mathbf{v} >,$$

where $\mathbf{v}$ is an arbitrary element from $\mathbf{V_0}$.

A commonly accepted way of the derivation of such estimates is exposed in the works of
M. Ainsworth, R. Becker, C. Johnson,
K. Eriksson, P. Hansbo, T. Oden,
S. Prhudome, R. Rannacher, E. Suli and other scientist, see, e.g.,
M. Ainsworth and J. T. Oden, *A posteriori error estimation in the finite element method,* Numer. Math., 60(1992), pp. 429-463.
W. Bangerth and R. Rannacher, *Adaptive finite element methods for differential equations,* Birkhäuser, Berlin, 2003.

It consists of the attraction the **adjoint problem**.

Adjoint problem

Find $\mathbf{u_a} \in \mathbf{V_0}$ such that

$$(\mathcal{A}^\star \mathbf{\Lambda u_a}, \mathbf{\Lambda w}) =< \ell, \mathbf{w} > \quad \forall \mathbf{w} \in \mathbf{V_0}, \tag{118}$$

where $\mathcal{A}^\star$ is the operator adjoint to $\mathcal{A}$, i.e.,

$$(\mathcal{A}\mathbf{y}, \mathbf{z}) = (\mathbf{y}, \mathcal{A}^\star \mathbf{z}) \quad \forall \mathbf{y}, \mathbf{z} \in \mathbf{U}.$$

Proposition

Let $\mathbf{u}$, $\mathbf{v}$ *and* $\mathbf{u_a}$, $\mathbf{v_a}$ *be exact and approximate solutions of the primal and adjoint problems respectively. Then,*

$$< \ell, \mathbf{u} - \mathbf{v} >:= \mathbf{E}(\mathbf{v}, \mathbf{v_a}) = \mathbf{E_0}(\mathbf{v}, \mathbf{v_a}) + \mathbf{E_1}(\mathbf{v}, \mathbf{v_a}), \tag{119}$$

where

$$\mathbf{E_0}(\mathbf{v}, \mathbf{v_a}) \ = < \mathbf{f}, \mathbf{v_a} > - (\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda v_a}) \tag{120}$$

and

$$\mathbf{E_1}(\mathbf{v}, \mathbf{v_a}) = (\mathcal{A}\mathbf{\Lambda}(\mathbf{u} - \mathbf{v}), \mathbf{\Lambda}(\mathbf{u_a} - \mathbf{v_a})). \tag{121}$$

Proof.

Since $\mathbf{u_a}$ is a solution of the adjoint problem, we have

$$
\begin{aligned}
< \ell, \mathbf{u} - \mathbf{v} > = \; & (\mathcal{A}^\star \mathbf{\Lambda u_a}, \mathbf{\Lambda}(\mathbf{u} - \mathbf{v})) = \\
& = (\mathcal{A}\mathbf{\Lambda}(\mathbf{u} - \mathbf{v}), \mathbf{\Lambda u_a}) = \\
& = (\mathcal{A}\mathbf{\Lambda}(\mathbf{u} - \mathbf{v}), \mathbf{\Lambda}(\mathbf{u_a} - \mathbf{v_a})) + \; < \mathbf{f}, \mathbf{v_a} > \; - (\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda v_a}).
\end{aligned}
$$

□

$\mathbf{E_0}(\mathbf{v}, \mathbf{v_a})$ is explicitly computable !
$\mathbf{E_1}(\mathbf{v}, \mathbf{v_a})$ contains unknown solutions of the two problems.
Let $\mathbf{V_h}$ and $\mathbf{V}_\tau$ be two finite-dimensional subspaces of $\mathbf{V_0}$, and let $\mathbf{v} = \mathbf{u_h}$ and $\mathbf{v_a} = \mathbf{u_{a\tau}}$, where $\mathbf{u_h}$ and $\mathbf{u_{a\tau}}$ are solutions of the problems

$$(\mathcal{A}\mathbf{\Lambda u_h}, \mathbf{\Lambda w_h}) =< \mathbf{f}, \mathbf{w_h} > \quad \forall \mathbf{w_h} \in \mathbf{V_h}, \tag{122}$$

$$(\mathcal{A}^\star \mathbf{\Lambda u_{a\tau}}, \mathbf{\Lambda w}_\tau) =< \ell, \mathbf{w}_\tau > \quad \forall \mathbf{w}_\tau \in \mathbf{V}_\tau. \tag{123}$$

In the particular case $\mathbf{V_h} \equiv \mathbf{V}_\tau$, the relation (122) implies that $\mathbf{E_0}(\mathbf{u_h}, \mathbf{u_{a\tau}}) = \mathbf{0}$ so that there remains the term containing the product of the (unknown) energy errors. On the contrary, for noncoinciding spaces both terms $\mathbf{E_0}$ and $\mathbf{E_1}$ are present. Moreover, it is easy to show that the term $\mathbf{E_0}$ dominates if $\mathbf{v_a}$ is close to $\mathbf{u_a}$. Indeed, if $\mathbf{v_a} \to \mathbf{u_a}$ in $\mathbf{V}$, then $\Lambda(\mathbf{u_a} - \mathbf{v_a}) \to \mathbf{0}$ in $\mathbf{U}$, so that

$$\mathbf{E_1}(\mathbf{v}, \mathbf{v_a}) \to \mathbf{0}.$$

However,

$$\mathbf{E_0}(\mathbf{v}, \mathbf{v_a}) + \mathbf{E_1}(\mathbf{v}, \mathbf{v_a}) =< \ell, \mathbf{u} - \mathbf{v} > \neq \mathbf{0}.$$

Hence, if $\mathbf{v_a}$ is sufficiently close to $\mathbf{u_a}$, then the directly computable term $\mathbf{E_0}(\mathbf{v}, \mathbf{v_a})$ contains the major part of the estimated quantity.

Key question: how to estimate

$$\mathbf{E}_1(\mathbf{v}, \mathbf{v_a}) = (\mathcal{A}\mathbf{\Lambda}(\mathbf{u} - \mathbf{v}), \mathbf{\Lambda}(\mathbf{u_a} - \mathbf{v_a})).$$

Two ways.

- Economical method that gives correct presentation on the error but does not give guaranteed error bounds,
- Computationally expensive method that provides two–sided guaranteed error bounds for the goal–oriented functional.

Economical way: use post-processing methods

Main idea: recover the unknown functions $\mathbf{\Lambda u}$ and $\mathbf{\Lambda u_a}$ by some post-processing techniques.

See: S. Korotov, P. Neittaanmaki and S. Repin,
A posteriori error estimation of goal-oriented quantities by the superconvergence patch recovery, J. Numer. Math. 11 (2003), 1, pp. 33-59.

Let G_h and G_τ be averaging operators defined on $\mathbf{V_h}$ and $\mathbf{V}_\tau$, respectively. Replace $\mathbf{E}(\mathbf{u_h}, \mathbf{u_{a\tau}})$ by the directly computable functional

$$\widetilde{E}(\mathbf{u_h}, \mathbf{u_{a\tau}}) := \mathbf{E_0}(\mathbf{u_h}, \mathbf{u_{a\tau}}) + \widetilde{\mathbf{E}}_\mathbf{1}(\mathbf{u_h}, \mathbf{u_{a\tau}}), \tag{124}$$

where

$$\widetilde{E}_1(\mathbf{u_h}, \mathbf{u_{a\tau}}) = (\mathcal{A}(\mathrm{G_h}(\mathbf{\Lambda u_h}) - \mathbf{\Lambda u_h}), \mathrm{G}_\tau(\mathbf{\Lambda u_{a\tau}}) - \mathbf{\Lambda u_{a\tau}}) .$$

If the operators G_h and G_τ perform a proper recovery of $\mathbf{\Lambda u_h}$ and $\mathbf{\Lambda u_{a\tau}}$, then it is natural to expect that the difference between $\mathbf{E_1}(\mathbf{u_h}, \mathbf{u_{a\tau}})$ and $\widetilde{E}(\mathbf{u_h}, \mathbf{u_{a\tau}})$ is given by higher order terms and, thus, the latter quantity can successfully be used instead of $\mathbf{E_1}$.

Two–sided estimates

Main idea: is to apply two-sided error estimates given by $\mathbf{M}_{\oplus}$ and $\mathbf{M}_{\ominus}$ to derive sharp bounds of $(\mathcal{A}\Lambda(\mathbf{u}-\mathbf{v}), \mathbf{\Lambda}(\mathbf{u_a}-\mathbf{v_a}))$ using a special representation of this term.

See

P. Neittaanmäki and S. Repin, *Reliable methods for computer simulation, Error control and a posteriori estimates*, Elsevier, New York, 2004.

Certainly, we may simply use the estimate

$$(\mathcal{A}\mathbf{\Lambda}(\mathbf{u}-\mathbf{v}),\mathbf{\Lambda}(\mathbf{u_a}-\mathbf{v_a}))\leq\|\!|\ \mathbf{\Lambda}(\mathbf{u}-\mathbf{v})\ |\!\|\!|\ \mathbf{\Lambda}(\mathbf{u_a}-\mathbf{v_a})\ |\!\|\ .$$

However, it may considerably **overestimate** the error.
Therefore, it is suggested to use the algebraic relation

$$\mathbf{2ab}=\left(\alpha\mathbf{a}+\frac{\mathbf{1}}{\alpha}\mathbf{b}\right)^{\mathbf{2}}-\alpha^{\mathbf{2}}\mathbf{a}^{\mathbf{2}}-\frac{\mathbf{1}}{\alpha^{\mathbf{2}}}\mathbf{b}^{\mathbf{2}}.$$

We have

$$\langle \ell, \mathbf{u} - \mathbf{v} \rangle := \mathbf{E}(\mathbf{v}, \mathbf{v_a}) = \mathbf{E_0}(\mathbf{v}, \mathbf{v_a}) + \mathbf{E_1}(\mathbf{v}, \mathbf{v_a}), \tag{125}$$

where $\mathbf{v}$ is an approximation of $\mathbf{u}$, $\mathbf{v_a}$ is an approximation of $\mathbf{u_a}$, and

$$\begin{aligned} \mathbf{E_0}(\mathbf{v}, \mathbf{v_a}) &= \langle \mathbf{f}, \mathbf{v_a} \rangle - (\mathcal{A}\mathbf{\Lambda}\mathbf{v}, \mathbf{\Lambda}\mathbf{v_a}), \\ \mathbf{E_1}(\mathbf{v}, \mathbf{v_a}) &= (\mathcal{A}\mathbf{\Lambda}\delta_\mathbf{f}, \mathbf{\Lambda}\delta_\ell), \quad \delta_f = (\mathbf{u} - \mathbf{v}),\ \delta_\ell = (\mathbf{u_a} - \mathbf{v_a}). \end{aligned}$$

By the algebraic identity

$$\mathbf{2}(\mathcal{A}\mathbf{\Lambda}\delta_\ell, \mathbf{\Lambda}\delta_\mathbf{f}) = \\ = \| \mathbf{\Lambda}(\alpha\delta_\mathbf{f} + \frac{\mathbf{1}}{\alpha}\delta_\ell) \|^2 - \alpha^2 \| \mathbf{\Lambda}\delta_\mathbf{f} \|^2 - \frac{\mathbf{1}}{\alpha^2} \| \mathbf{\Lambda}\delta_\ell \|^2 . \tag{126}$$

Assume that for the primal and adjoint problems we have found good upper and lower error bounds in the energy norm, i.e., we have $\mathbf{y}_\mathbf{f}^*$, $\mathbf{y}_\ell^*$, β_f, β_ℓ, $\mathbf{w}_\mathbf{f}$, and $\mathbf{w}_\ell$ such that

$$\mathbf{m}_\mathbf{f} := \boldsymbol{\mathcal{M}}_\ominus(\mathbf{v}, \mathbf{w}_\mathbf{f}) \le \frac{\mathbf{1}}{\mathbf{2}} \| \boldsymbol{\Lambda}\delta_\mathbf{f} \|^2 \le \boldsymbol{\mathcal{M}}_\oplus(\mathbf{v}, \beta_\mathbf{f}, \mathbf{y}_\mathbf{f}^*) := \mathbf{M}_\mathbf{f},$$

$$\mathbf{m}_\ell := \boldsymbol{\mathcal{M}}_\ominus(\mathbf{v}_\mathbf{a}, \mathbf{w}_\ell) \le \frac{\mathbf{1}}{\mathbf{2}} \| \boldsymbol{\Lambda}\delta_\ell \|^2 \le \boldsymbol{\mathcal{M}}_\oplus(\mathbf{v}_\mathbf{a}, \beta_\ell, \mathbf{y}_\ell^*) := \mathbf{M}_\ell.$$

Note that the quantity

$$\| \mathbf{\Lambda}(\alpha\delta_{\mathbf{f}} + \frac{\mathbf{1}}{\alpha}\delta_\ell) \|^2 = \| \mathbf{\Lambda}(\alpha\mathbf{u} + \frac{\mathbf{1}}{\alpha}\mathbf{u_a} - \alpha\mathbf{v} - \frac{\mathbf{1}}{\alpha}\mathbf{v_a}) \|^2$$

can be viewed as the norm of the difference between the exact solution $\mathbf{u}^\alpha_{\mathbf{f}\ell} \in \mathbf{V_0} + \big(\alpha + \frac{\mathbf{1}}{\alpha}\big)\mathbf{u_0}$ of the problem

$$(\mathcal{A}\mathbf{\Lambda u}^\alpha_{\mathbf{f}\ell}, \mathbf{\Lambda w}) + \langle \alpha\mathbf{f} + \frac{\mathbf{1}}{\alpha}\ell, \mathbf{w}\rangle = \mathbf{0}, \qquad \forall \mathbf{w} \in \mathbf{V_0}$$

and the function $\mathbf{v}^\alpha_{\mathbf{a_f}\ell} = \alpha\mathbf{v} + \frac{\mathbf{1}}{\alpha}\mathbf{v_a} \in \mathbf{V_0} + (\alpha + \frac{\mathbf{1}}{\alpha})\mathbf{u_0}$, which is an approximation of $\mathbf{u}^\alpha_{\mathbf{f}\ell}$.

To obtain two-sided bounds of $\| \mathbf{\Lambda}(\alpha\delta_{\mathbf{f}} + \frac{1}{\alpha}\delta_\ell) \|^2$ we use already known functions $\mathbf{y}^*_{\mathbf{f}}$, $\mathbf{y}^*_\ell$, $\mathbf{w}_{\mathbf{f}}$, and $\mathbf{w}_\ell$ and compute the numbers

$$\mathbf{m}_{\mathbf{f}\ell} = \boldsymbol{\mathcal{M}}_\ominus(\mathbf{u}^\alpha_{\mathbf{a}_{\mathbf{f}}\,\ell}, \alpha\mathbf{w}_{\mathbf{f}} + \frac{1}{\alpha}\mathbf{w}_\ell),$$

$$\mathbf{M}_{\mathbf{f}\ell} = \boldsymbol{\mathcal{M}}_\oplus(\mathbf{u}^\alpha_{\mathbf{a}_{\mathbf{f}}\ell}, \beta, \alpha\mathbf{y}^*_{\mathbf{f}} + \frac{1}{\alpha}\mathbf{y}^*_\ell).$$

We note that $\mathbf{m}_{\mathbf{f}\ell} = \mathbf{m}_{\mathbf{f}\ell}(\alpha)$, $\mathbf{M}_{\mathbf{f}\ell} = \mathbf{M}_{\mathbf{f}\ell}(\alpha, \beta)$ and positive numbers α and β can be taken arbitrary. By (126), we obtain

$$(\mathcal{A}\mathbf{\Lambda}\delta_{\mathbf{f}}, \mathbf{\Lambda}\delta_\ell) \le \inf_{\alpha,\beta\in\mathbb{R}_+} \{\mathbf{M}_{\mathbf{f}\,\ell} - \alpha^2\mathbf{m}_{\mathbf{f}} - \frac{\mathbf{m}_\ell}{\alpha^2}\} := \mathfrak{M}_\oplus,$$

$$(\mathcal{A}\mathbf{\Lambda}\delta_{\mathbf{f}}, \mathbf{\Lambda}\delta_\ell) \ge \sup_{\alpha\in\mathbb{R}_+} \{\mathbf{m}_{\mathbf{f}\ell} - \alpha^2\mathbf{M}_{\mathbf{f}} - \frac{\mathbf{M}_\ell}{\alpha^2}\} := \mathfrak{M}_\ominus.$$

Recalling (125), we now deduce a two-sided estimate

$$\mathfrak{M}_{\ominus} + \mathbf{E_0}(\mathbf{v}, \mathbf{v_a}) \le \langle \ell, \mathbf{u} - \mathbf{v} \rangle \le \mathfrak{M}_{\oplus} + \mathbf{E_0}(\mathbf{v}, \mathbf{v_a}), \tag{127}$$

If $\mathbf{y}_{\mathbf{f}}^*$, $\mathbf{y}_{\ell}^*$, $\mathbf{w}_{\mathbf{f}}$, and $\mathbf{w}_{\ell}$ provide accurate two-sided estimates of the energy error norms in the primal and adjoint problems, then $\mathbf{M}_{\mathbf{f}\,\ell}$ and $\mathbf{m}_{\mathbf{f}\,\ell}$ furnish accurate two-sided estimates for the norm $\| \mathbf{\Lambda}(\alpha\delta_{\mathbf{f}} + \frac{1}{\alpha}\delta_{\ell}) \|^2$ and, consequently, (127) yields sharp upper and lower bounds of $\langle \ell, \mathbf{u} - \mathbf{v} \rangle$.

Error estimates in terms of seminorms

Estimates of $\mathbf{u} - \mathbf{v}$ computed on a set of linear functionals can be used for an evaluation of this difference in some seminorms.
Let ω be a subdomain in Ω and $\{\varphi_1, \varphi_2, ..., \varphi_d\}$ be a set of functions that vanish in $\Omega \setminus \omega$.
By the method described above, we can estimate the quantities

$$\mathbf{I}_{\varphi_s} := \int_{\omega} \varphi_s(\mathbf{x})(\mathbf{u}(\mathbf{x}) - \mathbf{v}(\mathbf{x}))\, \mathbf{dx}.$$

Using these quantities we can estimate the error in terms of local seminorms

Let

$$\|\mathbf{u}-\mathbf{v}\|^2_{2,\omega} := \int_\omega |\mathbf{u}(\mathbf{x})-\mathbf{v}(\mathbf{x})|^2 \, \mathbf{dx}.$$

Note that

$$\|\mathbf{u}-\mathbf{v}\|_{2,\omega} = \sup_{\eta \in \mathbf{L^2}(\omega)} \frac{\int_\omega \eta(\mathbf{u}-\mathbf{v})\mathbf{dx}}{\|\eta\|_{2,\omega}},$$

and the supremum is attained if $\eta = \mathbf{u}-\mathbf{v}$.

Thus, if the difference $\mathbf{u} - \mathbf{v}$ is known to belong to a certain set $\boldsymbol{\Upsilon}(\omega) \subset \mathbf{L}^2(\omega)$, then the problem is reduced to the evaluation of the *seminorm*

$$|\mathbf{u} - \mathbf{v}|_{\boldsymbol{\Upsilon}} := \sup_{\eta \in \boldsymbol{\Upsilon}(\omega)} \frac{\int_\omega \eta(\mathbf{u} - \mathbf{v}) \mathbf{dx}}{\|\eta\|_{2,\omega}}.$$

In general, a seminorm does not provide complete information on the error, because the relation $|\mathbf{u} - \mathbf{v}|_{\boldsymbol{\Upsilon}} = \mathbf{0}$ does not mean that $\mathbf{u} = \mathbf{v}$. Nevertheless, seminorms may give a useful information if Υ contains sufficiently large amount of linearly independent functions.

Let Υ be a subspace made by φ_s, i.e.,

$$\Upsilon = \mathrm{Span}\left\{\varphi_1, \varphi_2, ..., \varphi_d\right\} = \left\{\sum_s \alpha_s \varphi_s\right\}.$$

Then,

$$|\mathbf{u}-\mathbf{v}|_{\Upsilon} = \sup_{\mathbf{a_i}\in\mathbb{R}} \frac{\sum_{\mathbf{i}=\mathbf{1}}^{\mathbf{d}} \mathbf{a_i} \int_{\omega} \varphi_{\mathbf{i}}(\mathbf{u}-\mathbf{v})\mathbf{dx}}{\left(\sum_{\mathbf{i,j}=\mathbf{1}}^{\mathbf{d}} \mathbf{a_i a_j} \int_{\omega} \varphi_{\mathbf{i}}\varphi_{\mathbf{j}}\mathbf{dx}\right)^{\frac{1}{2}}}. \tag{128}$$

This problem is equivalent to the following one:

$$\inf_{a_i \in \mathbb{R}} \sum_{i,j=1}^{d} a_i a_j \int_{\omega} \varphi_i \varphi_j dx.$$

for all a_i such that

$$\sum_{i=1}^{d} a_i \int_{\omega} \varphi_i (u - v) dx = 1$$

The latter problem can be reformulated as:

Problem. Find the quantity

$$\kappa^2(\omega, \boldsymbol{\Upsilon}) = \inf_{\mathbf{a} \in \mathbf{Q}} \mathbf{B}\mathbf{a} \cdot \mathbf{a}, \tag{129}$$

where $\mathbf{a} = (\mathbf{a_1}, \mathbf{a_2}, ..., \mathbf{a_d})$,

$$\mathbf{Q} = \{\mathbf{a} \in \mathbb{R}^{\mathbf{d}} \, | \mathbf{a} \cdot \mathbf{I} = \mathbf{1}\}, \qquad \mathbf{I} = (\mathbf{I}_{\varphi_1}, \mathbf{I}_{\varphi_2}, ..., \mathbf{I}_{\varphi_d}),$$

and

$$\mathbf{B} = \{\mathbf{b_{ij}}\}_{\mathbf{i,j=1,d}}, \quad \mathbf{b_{ij}} = \int_{\omega} \varphi_{\mathbf{i}} \varphi_{\mathbf{j}} \, \mathbf{dx}.$$

The problem (129) has a minimax form

$$\kappa^2(\omega, \boldsymbol{\Upsilon}) = \inf_{\mathbf{a}\in\mathbb{R}^{\mathbf{d}}} \sup_{\lambda\in\mathbb{R}} \{\mathbf{B}\ \mathbf{a}\cdot\mathbf{a} + \lambda[(\mathbf{a}\cdot\mathbf{I}) - \mathbf{1}]\} \geq$$

$$\geq \sup_{\lambda\in\mathbb{R}} \inf_{\mathbf{a}\in\mathbb{R}^d} \{\mathbf{B}\ \mathbf{a}\cdot\mathbf{a} + \lambda[(\mathbf{a}\cdot\mathbf{I}) - 1]\}. \quad (130)$$

Assume that the functions φ_i are chosen in such a way that $\mathbf{B}$ is a positive definite matrix. Then (130) holds as equality. The minimization problem that stands on the right-hand side in (130) has a solution

$$\mathbf{a} = -\frac{\lambda}{2}\mathbf{B}^{-1}\mathbf{I}.$$

Therefore,

$$\kappa^2(\omega, \Upsilon) = \sup_{\lambda \in \mathbb{R}} \{ -\frac{\lambda^2}{4} \mathbf{B}^{-1}\mathbf{l} \cdot \mathbf{l} - \lambda \} = \frac{1}{\mathbf{B}^{-1}\mathbf{l} \cdot \mathbf{l}},$$

and

$$|\mathbf{u} - \mathbf{v}|_{\Upsilon} = \frac{\mathbf{1}}{\kappa(\omega, \boldsymbol{\Upsilon})} = (\mathbf{B}^{-1}\mathbf{l} \cdot \mathbf{l})^{\frac{1}{2}}.$$

We see that the value of $|\mathbf{u} - \mathbf{v}|_{\Upsilon}$ is not difficult to find, provided that the functions $\{\varphi_s\}$ are properly chosen and the respective quantities l_{φ_s} are defined.

Let us show that if $\mathbf{v} \in \Upsilon(\omega)$ and $\mathbf{u}$ is a sufficiently smooth function, then **seminorm $|\mathbf{u}-\mathbf{v}|_\Upsilon$ gives a good estimate of $\|\mathbf{u}-\mathbf{v}\|_{2,\omega}$.** Let $\Upsilon(\omega)$ be a subspace of $L^2(\omega)$ and $\widehat{\mathbf{u}}$ be an arbitrary function from $\Upsilon(\omega)$. Then

$$\|\mathbf{u}-\mathbf{v}\|_{2,\omega} = \sup_{\eta \in \mathbf{L}^2(\omega)} \frac{\int_\omega ((\mathbf{u}-\widehat{\mathbf{u}})\eta + (\widehat{\mathbf{u}}-\mathbf{v})\eta)\, \mathbf{dx}}{\|\eta\|_{2,\omega}} \leq$$

$$\leq \sup_{\eta \in L^2(\omega)} \frac{\int_\omega (\widehat{\mathbf{u}}-\mathbf{v})\eta\, \mathbf{dx}}{\|\eta\|_{2,\omega}} + \|\mathbf{u}-\widehat{\mathbf{u}}\|_{2,\omega}.$$

Since $\widehat{\mathbf{u}} - \mathbf{v} \in \boldsymbol{\Upsilon}(\omega)$, we have

$$\|\mathbf{u} - \mathbf{v}\|_{2,\omega} \leq \sup_{\eta \in \boldsymbol{\Upsilon}(\omega)} \frac{\int_\omega (\widehat{\mathbf{u}} - \mathbf{v})\eta \, \mathbf{dx}}{\|\eta\|_{2,\omega}} + \|\mathbf{u} - \widehat{\mathbf{u}}\|_{2,\omega}.$$

By rearranging again the numerator of the fraction on the right-hand side, we arrive at the estimate

$$\|\mathbf{u} - \mathbf{v}\|_{2,\omega} = \sup_{\eta \in \boldsymbol{\Upsilon}(\omega)} \frac{\int_\omega ((\widehat{\mathbf{u}} - \mathbf{u})\eta + (\mathbf{u} - \mathbf{v})\eta) \, \mathbf{dx}}{\|\eta\|_{2,\omega}} + \|\mathbf{u} - \widehat{\mathbf{u}}\|_{2,\omega} \leq$$

$$\leq \sup_{\eta \in \Upsilon(\omega)} \frac{\int_\omega (\mathbf{u} - \mathbf{v})\eta \, \mathbf{dx}}{\|\eta\|_{2,\omega}} + 2\|\mathbf{u} - \widehat{\mathbf{u}}\|_{2,\omega} = |\mathbf{u} - \mathbf{v}|_{\boldsymbol{\Upsilon}} + 2\|\mathbf{u} - \widehat{\mathbf{u}}\|_{2,\omega},$$

in which $\widehat{\mathbf{u}}$ is an arbitrary function from $\Upsilon(\omega)$.

Hence, we see that

$$\|\mathbf{u}-\mathbf{v}\|_{2,\omega} \leq |\mathbf{u}-\mathbf{v}|_{\Upsilon} + 2\mathrm{dist}(\mathbf{u}, \boldsymbol{\Upsilon}(\omega)), \tag{131}$$

where

$$\mathrm{dist}(\mathbf{u}, \boldsymbol{\Upsilon}(\omega)) = \inf_{\widehat{\mathbf{u}} \in \boldsymbol{\Upsilon}(\omega)} \|\mathbf{u}-\widehat{\mathbf{u}}\|_{2,\omega}$$

is the distance between $\mathbf{u}$ and $\Upsilon(\omega)$.
If $\Upsilon(\Omega)$ is a set of polynomials and $\mathbf{u}$ is a smooth function then $\delta(\mathbf{u}, \boldsymbol{\Upsilon}(\omega))$ can be estimated with the help of well-known results of approximation theory.

By similar arguments, we obtain

$$\|\mathbf{u}-\mathbf{v}\|_{\mathbf{2},\omega} = \sup_{\eta\in\mathbf{L}^{\mathbf{2}}(\omega)} \frac{\int_{\omega}((\mathbf{u}-\widehat{\mathbf{u}})\eta+(\widehat{\mathbf{u}}-\mathbf{v})\eta)\,\mathbf{dx}}{\|\eta\|_{\mathbf{2},\omega}} \geq$$

$$\geq \sup_{\eta\in L^2(\omega)} \frac{\int_{\omega}(\widehat{\mathbf{u}}-\mathbf{v})\eta\mathbf{dx}}{\|\eta\|_{2,\omega}} - \|\mathbf{u}-\widehat{\mathbf{u}}\|_{\mathbf{2},\omega}.$$

Since $\widehat{\mathbf{u}}-\mathbf{v}$ belongs to $\Upsilon(\omega)$, we can replace in the first integral $\eta\in L^2$ by $\eta\in\Upsilon$.
Then, we obtain

$$\|\mathbf{u}-\mathbf{v}\|_{2,\omega} \geq \sup_{\eta\in\Upsilon(\omega)} \frac{\int_\omega ((\widehat{\mathbf{u}}-\mathbf{u})\eta + (\mathbf{u}-\mathbf{v})\eta)\ \mathbf{dx}}{\|\eta\|_{2,\omega}} - \|\mathbf{u}-\widehat{\mathbf{u}}\|_{2,\omega} \geq$$

$$\geq \sup_{\eta\in\Upsilon(\omega)} \frac{\int_\omega (\mathbf{u}-\mathbf{v})\eta\ \mathbf{dx}}{\|\eta\|_{2,\omega}} - \sup_{\eta\in L^2(\omega)} \frac{\int_\omega (\widehat{\mathbf{u}}-\mathbf{u})\eta\ \mathbf{dx}}{\|\eta\|_{2,\omega}} - \|\mathbf{u}-\widehat{\mathbf{u}}\|_{2,\omega} =$$

$$= |\mathbf{u}-\mathbf{v}|_\Upsilon - 2\|\mathbf{u}-\widehat{\mathbf{u}}\|_{2,\omega}, \quad \forall \widehat{\mathbf{u}} \in \Upsilon(\omega).$$

From here, it follows that

$$\|\mathbf{u}-\mathbf{v}\|_{2,\omega} \geq |\mathbf{u}-\mathbf{v}|_\Upsilon - 2\mathrm{dist}(\mathbf{u}, \Upsilon(\omega)). \tag{132}$$

By (131) and (132), we conclude that the error arising if $\|\mathbf{u}-\mathbf{u_h}\|_{2,\Omega}$ is replaced by $|\mathbf{u}-\mathbf{u_h}|_\Upsilon$ depends on the regularity of $\mathbf{u}$ and approximation properties of $\Upsilon(\omega)$.

Lecture 6

A POSTERIORI ESTIMATES FOR MIXED METHODS

Mixed approximations. A glance from the minimax theory

Consider our basic problem

$$\begin{aligned} &\mathbf{div A}\nabla \mathbf{u} + \mathbf{f} = \mathbf{0} \quad \text{in } \boldsymbol{\Omega}\,, \\ &\mathbf{u} = \mathbf{u_0} \text{on } \partial_1\boldsymbol{\Omega}\,, \\ &\mathbf{A}\nabla \mathbf{u}\cdot\mathbf{n} = \mathbf{F} \text{on } \partial_2\boldsymbol{\Omega}\,, \end{aligned}$$

$$\mathbf{c_1^2}|\xi|^2 \le \mathbf{A(x)}\xi\cdot\xi \le \mathbf{c_2^2}|\xi|^2 \qquad \forall \xi \in \mathbb{R}^{\mathbf{d}}\,, \text{ for a.e. } \mathbf{x}\in\boldsymbol{\Omega}\,,$$

where $\mathbf{u_0} \in \mathbf{H^1}(\boldsymbol{\Omega})$, $\mathbf{f} \in \mathbf{L_2}(\boldsymbol{\Omega})$, $\mathbf{F} \in \mathbf{L_2}(\partial_2\boldsymbol{\Omega})$. Functional spaces

$$\mathbf{V} := \mathbf{H^1}(\boldsymbol{\Omega})\,,\ \mathbf{V_0} := \{\mathbf{v}\in\mathbf{V} \mid \mathbf{v} = \mathbf{0} \text{ on } \partial_1\boldsymbol{\Omega}\}\,,\ \widehat{\mathbf{V}} := \mathbf{L_2}(\boldsymbol{\Omega})\,,$$

$$\mathbf{Q} := \mathbf{L_2}(\boldsymbol{\Omega};\mathbb{R}^{\mathbf{d}}) \qquad \widehat{\mathbf{Q}} := \mathbf{H}(\boldsymbol{\Omega};\mathbf{div})\,,$$

$$\widehat{\mathbf{Q}}^+ := \{\mathbf{y}\in\widehat{\mathbf{Q}} \mid \mathbf{y}\cdot\mathbf{n}\big|_{\partial_2\boldsymbol{\Omega}} \in \mathbf{L_2}(\partial_2\boldsymbol{\Omega})\}\,.$$

We recall that $\|\mathbf{q}\|_{\mathbf{div}}$ is the norm in $H(\Omega;\mathbf{div})$:

$$\|\mathbf{q}\|_{\mathbf{div}} := (\|\mathbf{q}\|^2 + \|\mathbf{div q}\|^2)^{1/2} \quad \forall \mathbf{q} \in \mathbf{Q}$$

and

$$\| \mathbf{q} \| := \left(\int_\Omega \mathbf{A q} \cdot \mathbf{q} \, \mathbf{dx} \right)^{1/2}, \quad \mathbf{q} \in \mathbf{Q}$$

$$\| \mathbf{q} \|_* := \left(\int_\Omega \mathbf{A}^{-1} \mathbf{q} \cdot \mathbf{q} \, \mathbf{dx} \right)^{1/2}$$

Note that,

$$\bar{\mathbf{c}}_1^2 |\xi|^2 \le \mathbf{A}^{-1}(\mathbf{x}) \xi \cdot \xi \le \bar{\mathbf{c}}_2^2 |\xi|^2 \qquad \forall \xi \in \mathbb{R}^{\mathbf{d}}, \text{ for a.e. } \mathbf{x} \in \mathbf{\Omega}$$

with $\bar{\mathbf{c}}_1 = \mathbf{1}/\mathbf{c}_2$, $\bar{\mathbf{c}}_2 = \mathbf{1}/\mathbf{c}_1$.

Generalized solution can be viewed as a saddle point of the Lagrangian

$$\mathbf{L}(\mathbf{v},\mathbf{q}) := \int_\Omega \left(\nabla \mathbf{v} \cdot \mathbf{q} - \frac{1}{2} \mathbf{A}^{-1} \mathbf{q} \cdot \mathbf{q} \right) \mathbf{dx} - \ell(\mathbf{v}) \,,$$

where $\ell(\mathbf{v}) = \int_\Omega \mathbf{f}\mathbf{v}\,\mathbf{dx} + \int_{\partial_2 \Omega} \mathbf{F}\mathbf{v}\,\mathbf{ds}$.

The problem of finding $(\mathbf{u},\mathbf{p}) \in \mathbf{V_0} + \mathbf{u_0} \times \mathbf{Q}$ such that

$$\mathbf{L}(\mathbf{u},\mathbf{q}) \;\le\; \mathbf{L}(\mathbf{u},\mathbf{p}) \;\le\; \mathbf{L}(\mathbf{v},\mathbf{p}) \quad \forall \mathbf{q} \in \mathbf{Q} \,,\; \forall \mathbf{v} \in \mathbf{V_0} + \mathbf{u_0} \tag{133}$$

leads to is the so-called **Primal Mixed Formulation**

In this formulation the **solution** of a PDE is understood as **the pair of functions** $(\mathbf{u}, \mathbf{p}) \in (\mathbf{V_0} + \mathbf{u_0}) \times \mathbf{Q}$ satisfying the relations

$$\int_\Omega \left(\mathbf{A}^{-1}\mathbf{p} - \nabla \mathbf{u} \right) \cdot \mathbf{q}\, \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{q} \in \mathbf{Q}\,, \tag{134}$$

$$\int_\Omega \mathbf{p} \cdot \nabla \mathbf{w}\, \mathbf{dx} - \ell(\mathbf{w}) = \mathbf{0} \qquad \forall \mathbf{w} \in \mathbf{V_0}\,. \tag{135}$$

In the PMM,

$\mathbf{p} = \mathbf{A}\nabla\mathbf{u}$, is satisfied in $\mathbf{L_2}(\Omega)$ − sense

$\mathbf{div p} + \mathbf{f} = \mathbf{0}$ in Ω and $\mathbf{p}\cdot\mathbf{n} = \mathbf{F}$ on $\partial_2\Omega$ are satisfied in a weak sense.

As we have seen in previous lectures $\mathbf{L}$ generates two functionals

$$\mathbf{J}(\mathbf{v}) := \sup_{\mathbf{q}\in\mathbf{Q}} \mathbf{L}(\mathbf{v},\mathbf{q}) = \frac{\mathbf{1}}{\mathbf{2}} \| \nabla \mathbf{v} \|^2 - \ell(\mathbf{v})$$

and

$$\mathbf{I}^*(\mathbf{q}) := -\frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{q} \|_*^2 - \ell(\mathbf{u_0}) + \int_\Omega \nabla \mathbf{u_0} \cdot \mathbf{q} \, \mathbf{dx} \, .$$

Also, we know that

$$\inf_{\mathbf{v}\in\mathbf{V_0}+\mathbf{u_0}} \mathbf{J}(\mathbf{v}) := \inf \mathcal{P} = \mathbf{L}(\mathbf{u},\mathbf{p}) = \sup \mathcal{P}^* := \sup_{\mathbf{q}\in\mathbf{Q}_\ell} \mathbf{I}^*(\mathbf{q}) \, , \qquad (136)$$

where $\mathbf{Q}_\ell := \{\mathbf{q} \in \mathbf{Q} \mid \int_\Omega \mathbf{q} \cdot \nabla \mathbf{w} \, \mathbf{dx} = \ell(\mathbf{w}) \ \forall \mathbf{w} \in \mathbf{V_0}\}$.

Primal Mixed Method (PMM)

Let $\mathbf{Q_h} \subset \mathbf{Q}$ and $\mathbf{V_{0h}} \subset \mathbf{V_0}$ are subspaces
constructed by FE approximation, then a discrete analog of (134)–(135) is the

Primal Mixed Finite Element Method .

See, e.g., F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*. Springer-Verlag, New York, 1991.

D. Braess. *Finite elements*. Cambridge University Press, Cambridge, 1997.

J. E. Roberts and J.-M. Thomas, *Mixed and Hybrid Methods*. In Handbook of Numerical Analysis, II, eds. P. G. Ciarlet and J.-L. Lions, North-Holland, Amsterdam, pp. 523–639, 1991.

In PMM, we need to find a pair of functions $(\mathbf{u_h}, \mathbf{p_h}) \in (\mathbf{V_{0h}} + \mathbf{u_0}) \times \mathbf{Q_h}$ such that

$$\int_\Omega (\mathbf{A}^{-1}\mathbf{p_h} - \nabla \mathbf{u_h}) \cdot \mathbf{q_h}\, \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{q_h} \in \mathbf{Q_h}\,, \tag{137}$$

$$\int_\Omega \mathbf{p_h} \cdot \nabla \mathbf{w_h}\, \mathbf{dx} - \ell(\mathbf{w_h}) = \mathbf{0} \quad \forall \mathbf{w_h} \in \mathbf{V_{0h}}\,. \tag{138}$$

In this formulation, $\mathbf{u_h}$ can be constructed by means of the Courant-type elements and $\mathbf{p_h}$ by piecewise constant functions.

Dual Mixed Method (DMM)

Another mixed formulation arises if we represent $\mathbf{L}$ in a somewhat different form. First, we introduce the functional $\mathbf{g} : (\mathbf{V_0} + \mathbf{u_0}) \times \widehat{\mathbf{Q}} \to \mathbb{R}$ by the relation

$$\mathbf{g(v,q)} := \int_\Omega (\nabla \mathbf{v} \cdot \mathbf{q} + \mathbf{v}(\mathbf{divq}))\, \mathbf{dx}\,.$$

We have

$$\mathbf{L(v,q)} = \int_\Omega \left(\nabla \mathbf{v} \cdot \mathbf{q} - \frac{1}{2} \mathbf{A}^{-1} \mathbf{q} \cdot \mathbf{q} \right) \mathbf{dx} - \ell(\mathbf{v}) =$$

$$= \mathbf{g(v,q)} - \int_\Omega \mathbf{v}(\mathbf{divq})\, \mathbf{dx} - \frac{1}{2} \parallel \mathbf{q} \parallel_*^2 - \ell(\mathbf{v})\,.$$

Introduce the set

$$\widehat{\mathbf{Q}}_{\mathbf{F}} := \{\mathbf{q} \in \widehat{\mathbf{Q}} \mid \mathbf{g}(\mathbf{w}, \mathbf{q}) = \int_{\partial_2 \Omega} \mathbf{F}\mathbf{w}\, \mathbf{ds} \;\; \forall \mathbf{w} \in \mathbf{V_0}\}\,.$$

Note that for $\mathbf{q} \in \widehat{\mathbf{Q}}_{\mathbf{F}}$ we have

$$\mathbf{g}(\mathbf{v}, \mathbf{q}) = \mathbf{g}(\mathbf{w} + \mathbf{u_0}, \mathbf{q}) = \mathbf{g}(\mathbf{w}, \mathbf{q}) + \mathbf{g}(\mathbf{u_0}, \mathbf{q}) =$$

$$= \int_{\partial_2 \Omega} \mathbf{F}\mathbf{w}\, \mathbf{ds} + \mathbf{g}(\mathbf{u_0}, \mathbf{q}) \;\forall \mathbf{w} \in \mathbf{V_0}\,.$$

Therefore, if the variable $\mathbf{q}$ is taken not from $\mathbf{Q}$ but from the narrower set $\widehat{\mathbf{Q}}_F$, then the Lagrangian can be written as

$$\widehat{\mathbf{L}}(\mathbf{v}, \mathbf{q}) :=$$

$$-\frac{\mathbf{1}}{\mathbf{2}} \parallel\!\!| \, \mathbf{q} \, |\!\!\parallel_*^2 - \int_{\Omega} \mathbf{v}(\mathbf{divq})\, \mathbf{dx} - \int_{\Omega} \mathbf{fv}\, \mathbf{dx} - \int_{\partial_2 \Omega} \mathbf{F}\mathbf{u_0}\, \mathbf{ds} + \mathbf{g}(\mathbf{u_0}, \mathbf{q})\,.$$

We observe Note the new Lagrangian $\widehat{\mathbf{L}}$
is defined on a wider set of primal functions $\mathbf{v} \in \widehat{\mathbf{V}}$, but uses a narrower set $\widehat{\mathbf{Q}}_F$ for the fluxes.

The problem of finding $(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) \in \widehat{\mathbf{V}} \times \widehat{\mathbf{Q}}_F$ such that

$$\widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{q}}) \leq \widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) \leq \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \widehat{\mathbf{p}}) \quad \forall \widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}\,, \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}} \tag{139}$$

lead to is the so-called

Dual Mixed Formulation

of the problem in question (see, e.g., F. Brezzi and M. Fortin).

From (139) we obtain the necessary conditions for the dual mixed formulation. Since

$$\widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{q}}) \leq \widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) \quad \forall \widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{F}},$$

we have

$$-\frac{\mathbf{1}}{\mathbf{2}} \parallel\!\!| \widehat{\mathbf{p}} + \lambda\boldsymbol{\eta} \parallel\!\!|_*^2 - \int_{\Omega} \widehat{\mathbf{u}}(\mathbf{div}(\widehat{\mathbf{p}} + \lambda\boldsymbol{\eta}) - \mathbf{f}\widehat{\mathbf{u}})\mathbf{dx} - \int_{\partial_2\Omega} \mathbf{F}\mathbf{u_0}\,\mathbf{ds} + \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{p}} + \lambda\boldsymbol{\eta}) \leq$$

$$-\frac{\mathbf{1}}{\mathbf{2}} \parallel\!\!| \widehat{\mathbf{p}} \parallel\!\!|_*^2 - \int_{\Omega} \widehat{\mathbf{u}}(\mathbf{div}\widehat{\mathbf{p}})\,\mathbf{dx} - \int_{\Omega} \mathbf{f}\widehat{\mathbf{u}}\,\mathbf{dx} - \int_{\partial_2\Omega} \mathbf{F}\mathbf{u_0}\,\mathbf{ds} + \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{p}}),$$

where λ is a real number and $\boldsymbol{\eta}$ is a function in $\widehat{\mathbf{Q}}_0 := \widehat{\mathbf{Q}}_F$ with $\mathbf{F} = \mathbf{0}$. Now, arrive at the relation

$$-\lambda \int_{\Omega} (\mathbf{A}^{-1}\widehat{\mathbf{p}} \cdot \boldsymbol{\eta} + \widehat{\mathbf{u}}(\mathbf{div}\boldsymbol{\eta}))\mathbf{dx} + \lambda\mathbf{g}(\mathbf{u_0}, \boldsymbol{\eta}) \leq \frac{\lambda^2}{\mathbf{2}} \int_{\Omega} \mathbf{A}^{-1}\boldsymbol{\eta} \cdot \boldsymbol{\eta}\mathbf{dx}.$$

Rewrite it as

$$\int_\Omega (\mathbf{A}^{-1}\widehat{\mathbf{p}} \cdot \boldsymbol{\eta} + \widehat{\mathbf{u}}(\mathbf{div}\boldsymbol{\eta}))\mathbf{dx} - \mathbf{g}(\mathbf{u_0}, \boldsymbol{\eta}) \geq \frac{\lambda}{\mathbf{2}} \int_\Omega \mathbf{A}^{-1}\boldsymbol{\eta} \cdot \boldsymbol{\eta}\mathbf{dx}.$$

Since $\lambda > 0$ can be taken arbitrarily small, the latter relation may hold only if

$$\int_\Omega (\mathbf{A}^{-1}\widehat{\mathbf{p}} \cdot \boldsymbol{\eta} + \widehat{\mathbf{u}}\mathbf{div}\boldsymbol{\eta})\mathbf{dx} - \mathbf{g}(\mathbf{u_0}, \boldsymbol{\eta}) \geq \mathbf{0}.$$

But $\boldsymbol{\eta}$ is an arbitrary element of a linear manifold $\widehat{\mathbf{Q}}_\mathbf{0}$, so that $+\boldsymbol{\eta}$ can be replaced by $-\boldsymbol{\eta}$ what leads to the conclusion that

$$\int_\Omega (\mathbf{A}^{-1}\widehat{\mathbf{p}} \cdot \boldsymbol{\eta} + \widehat{\mathbf{u}}\mathbf{div}\boldsymbol{\eta})\mathbf{dx} - \mathbf{g}(\mathbf{u_0}, \boldsymbol{\eta}) = \mathbf{0} \quad \forall \boldsymbol{\eta} \in \widehat{\mathbf{Q}}_\mathbf{0}.$$

From

$$\widehat{\mathbf{L}}(\widehat{\mathbf{u}},\widehat{\mathbf{p}}) \;\leq\; \widehat{\mathbf{L}}(\widehat{\mathbf{u}}+\widehat{\mathbf{v}},\widehat{\mathbf{p}}) \quad \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}} := \mathbf{L}^2(\Omega)$$

we observe that the terms of $\widehat{\mathbf{L}}$ linear with respect to the "pressure" must vanish. Namely, we obtain

$$\int_\Omega (\widehat{\mathbf{v}}\mathbf{div}\widehat{\mathbf{p}} + \mathbf{f}\widehat{\mathbf{v}})\mathbf{dx} = \mathbf{0}$$

Thus, we arrive at the system

$$\int_\Omega \left(\mathbf{A}^{-1}\widehat{\mathbf{p}}\cdot\widehat{\mathbf{q}} + (\mathbf{div}\widehat{\mathbf{q}})\widehat{\mathbf{u}}\right)\,\mathbf{dx} = \mathbf{g}(\mathbf{u_0},\widehat{\mathbf{q}}) \quad \forall \widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_0\,, \tag{140}$$

$$\int_\Omega (\mathbf{div}\widehat{\mathbf{p}} + \mathbf{f})\widehat{\mathbf{v}}\,\mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}}\,. \tag{141}$$

We observe that now the condition

$$\mathbf{div}\widehat{\mathbf{p}} + \mathbf{f} = \mathbf{0}$$

is satisfied in a "strong" ($\mathbf{L_2}$) sense, the Neumann type boundary condition is viewed as the essential boundary condition, and the relation

$$\widehat{\mathbf{p}} = \mathbf{A}\nabla\widehat{\mathbf{u}}$$

and the Dirichlet type boundary condition are satisfied in a weak sense. These properties of the DMM lead to that the respective finite dimensional formulations are better adapted to the satisfaction of the equilibrium type relations for the fluxes. This fact is important in many applications where a sharp satisfaction of the equilibrium relations is required.

The Lagrangian $\widehat{\mathbf{L}}$ also generates two functionals

$$\widehat{\mathbf{J}}(\widehat{\mathbf{v}}) := \sup_{\widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}} \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \widehat{\mathbf{q}}) \quad \text{and} \quad \widehat{\mathbf{I}}^*(\widehat{\mathbf{q}}) := \inf_{\widehat{\mathbf{v}} \in \widehat{\mathbf{V}}} \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \widehat{\mathbf{q}}) \,.$$

The two corresponding variational problems are

$$\inf_{\widehat{\mathbf{v}} \in \widehat{\mathbf{V}}} \widehat{\mathbf{J}}(\widehat{\mathbf{v}}) \quad \text{and} \quad \sup_{\widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}} \widehat{\mathbf{I}}^*(\widehat{\mathbf{q}}).$$

They are called Problems $\widehat{\mathcal{P}}$ and $\widehat{\mathcal{P}}^*$, respectively. Note that the functional $\widehat{\mathbf{J}}$ (unlike $\mathbf{J}$) has no simple explicit form. However, we can prove the solvability of Problem $\widehat{\mathcal{P}}$ by the following Lemma.

Lemma

For any $\widehat{\mathbf{v}} \in \widehat{\mathbf{V}}$ *and* $\mathbf{F} \in \mathbf{L}_2(\partial_2\Omega)$ *there exists* $\mathbf{p}^{\mathbf{v}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}$ *such that*

$$\mathbf{div}\mathbf{p}^{\mathbf{v}} + \widehat{\mathbf{v}} = \mathbf{0} \ \text{ in } \Omega\,, \tag{142}$$

$$\| \mathbf{p}^{\mathbf{v}} \|_* \leq \mathbf{C}_\Omega \left(\|\widehat{\mathbf{v}}\| + \|\mathbf{F}\|_{\partial_2\Omega} \right) . \tag{143}$$

Proof. We know that the boundary-value problem

$$\begin{aligned} \mathbf{div}\mathbf{A}\nabla\mathbf{u}^{\mathbf{v}} + \widehat{\mathbf{v}} = \mathbf{0} \quad & \text{in } \Omega\,, \\ \mathbf{u}^{\mathbf{v}} = \mathbf{0} \quad & \text{on } \partial_1\Omega\,, \\ \mathbf{A}\nabla\mathbf{u}^{\mathbf{v}} \cdot \mathbf{n} = \mathbf{F} \quad & \text{on } \partial_2\Omega \end{aligned}$$

possesses the unique solution $\mathbf{u}^{\mathbf{v}} \in \mathbf{V_0}$.

For it and the energy estimate

$$\| \nabla \mathbf{u}^{\mathbf{v}} \| \leq \mathbf{C}_{\Omega} \left(\|\widehat{\mathbf{v}}\| + \|\mathbf{F}\|_{\partial_2 \Omega} \right)$$

holds. Let $\mathbf{p}^{\mathbf{v}} := \mathbf{A} \nabla \mathbf{u}^{\mathbf{v}}$. We have

$$\mathbf{div} \mathbf{p}^{\mathbf{v}} + \widehat{\mathbf{v}} = \mathbf{0}.$$

Obviously, $\mathbf{p}^{\mathbf{v}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}$ and, since

$$\| \mathbf{p}^{\mathbf{v}} \|_*^2 = \int_{\Omega} \mathbf{A}^{-1} (\mathbf{A} \nabla \mathbf{u}^{\mathbf{v}}) \cdot (\mathbf{A} \nabla \mathbf{u}^{\mathbf{v}}) \, \mathbf{dx} = \| \nabla \mathbf{u}^{\mathbf{v}} \|^2 ,$$

we find that (143) also holds.
□

By the Lemma we can easily prove the coercivity of $\widehat{\mathbf{J}}$ on $\widehat{\mathbf{V}}$. Indeed,

$$\widehat{\mathbf{J}}(\widehat{\mathbf{v}}) \geq \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \alpha \mathbf{p}^{\mathbf{v}}) =$$

$$-\frac{1}{2} \| \alpha \mathbf{p}^{\mathbf{v}} \|_{*}^{2} - \alpha \int_{\Omega} \widehat{\mathbf{v}} (\mathbf{div} \mathbf{p}^{\mathbf{v}}) \, \mathbf{dx} - \int_{\Omega} \mathbf{f} \widehat{\mathbf{v}} \, \mathbf{dx} - \int_{\partial_2 \Omega} \mathbf{F} \mathbf{u_0} \, \mathbf{ds} + \mathbf{g}(\mathbf{u_0}, \alpha \mathbf{p}^{\mathbf{v}}) =$$

$$= -\frac{1}{2} \alpha^2 \| \mathbf{p}^{\mathbf{v}} \|_{*}^{2} + \alpha \|\widehat{\mathbf{v}}\|^2 - \|\mathbf{f}\| \|\widehat{\mathbf{v}}\| + \mathbf{g}(\mathbf{u_0}, \alpha \mathbf{p}^{\mathbf{v}}) - \int_{\partial_2 \Omega} \mathbf{F} \mathbf{u_0} \, \mathbf{ds} \, .$$

Here $|\mathbf{g}(\mathbf{u_0}, \alpha \mathbf{p}^{\mathbf{v}})| \leq \alpha \|\mathbf{p}^{\mathbf{v}}\|_{\mathbf{div}} \|\mathbf{u_0}\|_{\mathbf{1,2,\Omega}}$ and

$$\|\mathbf{p}^{\mathbf{v}}\|_{\mathbf{div}}^2 = \|\mathbf{p}^{\mathbf{v}}\|^2 + \|\mathbf{div} \mathbf{p}^{\mathbf{v}}\|^2 \leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_{\mathbf{1}}} \| \mathbf{p}^{\mathbf{v}} \|_{*}^{2} + \|\widehat{\mathbf{v}}\|^2 \leq$$

$$\leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_{\mathbf{1}}} \mathbf{C}_{\Omega}^{2} \left(\|\widehat{\mathbf{v}}\| + \|\mathbf{F}\|_{\partial_2 \Omega} \right)^2 + \|\widehat{\mathbf{v}}\|^2 \, .$$

Therefore

$$\widehat{\mathbf{J}}(\widehat{\mathbf{v}}) \geq -\frac{\mathbf{1}}{\mathbf{2}}\alpha^2 \mathbf{C}_{\Omega}^2 \|\widehat{\mathbf{v}}\|^2 + \alpha\|\widehat{\mathbf{v}}\|^2 + \boldsymbol{\Theta}(\|\widehat{\mathbf{v}}\|) + \boldsymbol{\Theta}_{\mathbf{0}}\,,$$

where $\boldsymbol{\Theta}(\|\widehat{\mathbf{v}}\|)$ contains the terms linear with respect to $\|\widehat{\mathbf{v}}\|$ and $\boldsymbol{\Theta}_{\mathbf{0}}$ does not depend on $\widehat{\mathbf{v}}$. Take $\alpha = \mathbf{1}/\mathbf{C}_{\Omega}^2$. Then

$$\widehat{\mathbf{J}}(\widehat{\mathbf{v}}) \geq \frac{\mathbf{1}}{\mathbf{2C}_{\Omega}^2}\|\widehat{\mathbf{v}}\|^2 + \boldsymbol{\Theta}(\|\widehat{\mathbf{v}}\|) + \boldsymbol{\Theta}_{\mathbf{0}} \longrightarrow +\infty \text{ as } \|\widehat{\mathbf{v}}\| \to \infty\,.$$

It is not difficult to prove that the functional $\widehat{J}$ is convex and lower semicontinuous. Therefore, Problem $\widehat{\mathcal{P}}$ has a solution $\widehat{\mathbf{u}}$.

Inf-Sup condition for the dual mixed formulation

Corollary

Lemma implies the *inf-sup* condition

$$\inf_{\substack{\phi\in \mathbf{L^2(\Omega)} \\ \psi\in \mathbf{L^2(\partial_2\Omega)}}} \sup_{\mathbf{q}\in \widehat{\mathbf{Q}}_\mathbf{F}} \frac{\int_\Omega \phi \mathbf{div q}\, \mathbf{dx} + \int_{\partial_2\Omega} \psi \mathbf{q}\cdot\mathbf{n}\, \mathbf{ds}}{\|\mathbf{q}\|_{\mathbf{div}} \left(\|\phi\|^2 + \|\psi\|^2_{\partial_2\Omega}\right)^{1/2}} \geq \mathbf{C_0} > \mathbf{0}\,.$$

The Dual Problem with respect to the Lagrangian $\widehat{\mathsf{L}}$

Let us now construct the dual functional $\widehat{\mathsf{I}}^*$. It is easy to see that

$$\widehat{\mathsf{I}}^*(\widehat{\mathbf{q}}) = \inf_{\widehat{\mathbf{v}}} \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \widehat{\mathbf{q}}) =$$

$$= \inf_{\widehat{\mathbf{v}}} \left\{ -\frac{1}{2} \| \widehat{\mathbf{q}} \|_*^2 - \int_\Omega \mathbf{v}(\mathbf{div}\widehat{\mathbf{q}})\mathbf{dx} - \int_\Omega \mathbf{f}\mathbf{v}\mathbf{dx} - \int_{\partial_2\Omega} \mathbf{F}\mathbf{u_0}\mathbf{ds} + \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{q}}) \right\} =$$

$$= -\frac{1}{2} \| \widehat{\mathbf{q}} \|_*^2 + \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{q}}) - \int_{\partial_2\Omega} \mathbf{F}\mathbf{u_0}\, \mathbf{ds}$$

provided that $\mathbf{div}\widehat{\mathbf{q}} + \mathbf{f} = \mathbf{0}$ (in the $\mathbf{L_2}$-sense). In all other cases $\widehat{\mathsf{I}}^*(\widehat{\mathbf{q}}) = -\infty$.

Recalling that $\mathbf{div}\widehat{\mathbf{q}} = -\mathbf{f}$ (in $L_2(\Omega)$-sense), we find that the dual functional for such a case has the form

$$\widehat{\mathbf{I}}^*(\mathbf{q}) = -\frac{\mathbf{1}}{\mathbf{2}} \parallel \widehat{\mathbf{q}} \parallel_*^2 + \int_{\Omega} (\nabla \mathbf{u_0} \cdot \widehat{\mathbf{q}} - \mathbf{f u_0})\, \mathbf{dx} - \int_{\partial_2 \Omega} \mathbf{F u_0}\, \mathbf{ds}$$

$$= \int_{\Omega} \nabla \mathbf{u_0} \cdot \widehat{\mathbf{q}} \mathbf{dx} - \frac{\mathbf{1}}{\mathbf{2}} \parallel \widehat{\mathbf{q}} \parallel_*^2 - \ell(\mathbf{u_0})\,,$$

Since $\widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}$, we have

$$\int_{\Omega} \nabla \mathbf{w} \cdot \widehat{\mathbf{q}} \mathbf{dx} = -\int_{\Omega} (\mathbf{div}\widehat{\mathbf{q}}) \mathbf{w} \mathbf{dx} + \int_{\partial_2 \Omega} \mathbf{F w}\, \mathbf{ds} \quad \forall \mathbf{w} \in \mathbf{V_0}\,.$$

we see that $\widehat{\mathbf{q}}$ satisfies the relation

$$\int_{\Omega} \nabla \mathbf{w} \cdot \widehat{\mathbf{q}} \mathbf{dx} = \ell(\mathbf{w}) \quad \forall \mathbf{w} \in \mathbf{V_0}\,.$$

In other cases, $\widehat{\mathbf{I}}^*(\widehat{\mathbf{q}}) = -\infty$.

Thus, Problems $\mathcal{P}^*$ and $\widehat{\mathcal{P}}^*$ coincide and are reduced to the maximization of I^* on the set Q_ℓ. This means that

$$\sup \mathcal{P}^* = \sup \widehat{\mathcal{P}}^*.$$

Since the saddle point of $\widehat{\mathsf{L}}$ exists, we have

$$\widehat{\mathsf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) = \inf \widehat{\mathcal{P}} = \sup \widehat{\mathcal{P}}^*,$$

but

$$\sup \widehat{\mathcal{P}}^* = \sup \mathcal{P}^* = \inf \mathcal{P}.$$

Thus, we infer that

$$\inf \widehat{\mathcal{P}} = \inf \mathcal{P}.$$

Thus, we conclude that $\mathbf{u} \in \mathbf{V_0} + \mathbf{u_0}$ (minimizer of $\mathcal{P}$) also minimizes $\widehat{\mathbf{J}}$ on $\widehat{\mathbf{V}}$.
Analogously, if $\mathbf{p} \in \mathbf{Q}_\ell$ is the maximizer of Problem $\mathcal{P}^*$, then

$$\int_\Omega \nabla \mathbf{w} \cdot \mathbf{p}\, \mathbf{dx} = \int_\Omega \mathbf{f}\mathbf{w}\, \mathbf{dx} + \int_{\partial_2 \Omega} \mathbf{F}\mathbf{w}\, \mathbf{ds} \quad \forall \mathbf{w} \in \mathbf{V_0}\,.$$

From here we see that $\mathbf{divp} + \mathbf{f} = \mathbf{0}$ a.e. in Ω and, hence,

$$\int_\Omega (\nabla \mathbf{w} \cdot \mathbf{p} + (\mathbf{divp})\mathbf{w})\, \mathbf{dx} = \int_{\partial_2 \Omega} \mathbf{F}\mathbf{w}\, \mathbf{ds} \quad \forall \mathbf{w} \in \mathbf{V_0}\,,$$

that is $\mathbf{p} \in \widehat{\mathbf{Q}}_\mathbf{F}$. Thus, $\mathbf{p}$ is also the maximizer of Problem $\widehat{\mathcal{P}}^*$.
The reverse statement that the solutions of $\widehat{\mathcal{P}}$, $\widehat{\mathcal{P}}^*$ are also the solutions of $\mathcal{P}$, $\mathcal{P}^*$ is not difficult to prove as well.

Hence, both mixed formulations have the same solution $(\mathbf{u}, \mathbf{p})$ which is in fact the generalized solution of our problem.

Finite dimensional formulations

Let

$$\widehat{\mathbf{V}}_h \subset \widehat{\mathbf{V}}, \qquad \widehat{\mathbf{Q}}_{0h} \subset \widehat{\mathbf{Q}}_0 \qquad \widehat{\mathbf{Q}}_{Fh} \subset \widehat{\mathbf{Q}}_F$$

A discrete analog of the dual mixed formulation is: Find $(\widehat{\mathbf{u}}_h, \widehat{\mathbf{p}}_h) \in \widehat{\mathbf{V}}_h \times \widehat{\mathbf{Q}}_{Fh}$ such that

$$\int_\Omega \left(\mathbf{A}^{-1} \widehat{\mathbf{p}}_\mathbf{h} \cdot \widehat{\mathbf{q}}_\mathbf{h} + \widehat{\mathbf{u}}_\mathbf{h} \mathbf{div} \widehat{\mathbf{q}}_\mathbf{h} \right) \mathbf{dx} = \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{q}}_\mathbf{h}) \quad \forall \widehat{\mathbf{q}}_\mathbf{h} \in \widehat{\mathbf{Q}}_{\mathbf{0h}}, \tag{144}$$

$$\int_\Omega (\mathbf{div} \widehat{\mathbf{p}}_\mathbf{h} + \mathbf{f}) \widehat{\mathbf{v}}_\mathbf{h} \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}}_\mathbf{h} \in \widehat{\mathbf{V}}_\mathbf{h}. \tag{145}$$

Error analysis for DMM

First we will obtain a priori error estimates for the dual mixed method and after that we will derive computable upper bounds for the quantities

$$\| \nabla(\mathbf{u} - \mathbf{u_h}) \| , \ \| \mathbf{p} - \mathbf{p_h} \|_* , \ \|\mathbf{p} - \widehat{\mathbf{p}}_\mathbf{h}\|_{\mathbf{div}} .$$

A priori error estimates for DMM

Below we will show a simple way of the derivation of projection type error estimates for the dual mixed method. By combining them with standard interpolation results, one can obtain known rate convergence estimates. A detailed exposition of this subject can be found in the above cited books. Here, we present a simplified version, which, however contains the main ideas of the a priori error analysis for the dual mixed approximations.

For the sake of simplicity we will consider the case of uniform Dirichlét boundary conditions and a constant matrix $\mathbf{A}$. In this case, the basic system is as follows

$$\int_\Omega \left(\mathbf{A}^{-1}\widehat{\mathbf{p}}\cdot\widehat{\mathbf{q}} + (\mathbf{div}\widehat{\mathbf{q}})\widehat{\mathbf{u}}\right)\,\mathbf{dx} = \mathbf{0} \quad \forall\widehat{\mathbf{q}}\in\widehat{\mathbf{Q}}_0\,,$$

$$\int_\Omega (\mathbf{div}\widehat{\mathbf{p}} + \mathbf{f})\widehat{\mathbf{v}}\,\mathbf{dx} = \mathbf{0} \quad \forall\widehat{\mathbf{v}}\in\widehat{\mathbf{V}}\,.$$

Since there is no Neumann part of the boundary, $\widehat{\mathbf{Q}}_F$ and $\widehat{\mathbf{Q}}_0$ coincides with $\widehat{\mathbf{Q}} := \mathbf{H}(\Omega, \mathbf{div})$.

In the considered, case the system of DMM is as follows

$$\int_{\Omega} \left(\mathbf{A}^{-1} \widehat{\mathbf{p}}_h \cdot \widehat{\mathbf{q}}_h + \widehat{\mathbf{u}}_h \mathbf{div} \widehat{\mathbf{q}}_h \right) \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_h \in \widehat{\mathbf{Q}}_h,$$

$$\int_{\Omega} (\mathbf{div} \widehat{\mathbf{p}}_h + \mathbf{f}) \widehat{\mathbf{v}}_h \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}}_h \in \widehat{\mathbf{V}}_h.$$

Assumptions.

(a) $\mathcal{T}_h$ is a regular triangulation of a polygonal domain Ω.

(b) $\widehat{\mathbf{V}}_h = \{\mathbf{v_h} \in \mathbf{L^2} \,|\, \mathbf{v_h} \in \mathbf{P^0(T)} \quad \forall \mathbf{T} \in \mathcal{T}_\mathbf{h}\}$.

(c) $\widehat{\mathbf{Q}}_h = \{\mathbf{q_h} \in \mathbf{H(\Omega, div)} \,|\, \mathbf{q_h} \in \mathbf{RT^0(T)} \; \forall \mathbf{T} \in \mathcal{T}_\mathbf{h}\}$.

(d) $\mathbf{f} \in \mathbf{P^0(T)}, \quad \forall \mathbf{T} \in \mathcal{T}_\mathbf{h}$

Note that under the assumptions made

$$\mathbf{div p_h} + \mathbf{f} = \mathbf{0} \qquad \text{on any } \mathbf{T}.$$

Indeed, this fact directly follows from the relation

$$\int_{\Omega} (\mathbf{div}\widehat{\mathbf{p}}_\mathbf{h} + \mathbf{f})\widehat{\mathbf{v}}_\mathbf{h} \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}}_\mathbf{h} \in \widehat{\mathbf{V}}_\mathbf{h}.$$

Therefore $\mathbf{p_h} \in \mathbf{Q_f}$.

Compatibility and stability conditions

In order to provide the stability of the discrete DM formulation we need additional assumptions.
We assume that a pair of finite dimensional spaces $\widehat{\mathbf{V}}_\mathbf{h}$, $\widehat{\mathbf{Q}}_\mathbf{h}$ satisfies the following condition:

For any $\mathbf{v_h} \in \widehat{\mathbf{V}}_\mathbf{h}$ exists $\mathbf{q}^\mathbf{v}_\mathbf{h} \in \widehat{\mathbf{Q}}_\mathbf{h}$ such that

$$\mathbf{div}\mathbf{q}^\mathbf{v}_\mathbf{h} = \mathbf{v_h} \quad \text{(compatibility)}, \tag{146}$$

$$\|\mathbf{q}^\mathbf{v}_\mathbf{h}\| \leq \mathbf{C}\|\mathbf{v_h}\| \quad \text{(stability)}. \tag{147}$$

We will show that the above two conditions are the sufficient conditions for proving that discrete DM problem is
(a) correct (e.g., has a solution),
(b) stable,
(c) has a projection type error estimate.

Discrete Inf-Sup condition

From (146) and (147), it follows that

$$\inf_{v_h \in \widehat{V}_h} \sup_{q_h \in \widehat{Q}_h} \frac{\int_\Omega v_h \, \mathrm{div} q_h \, dx}{\|v_h\| \|q_h\|_{\mathrm{div}}} \geq \mathbb{C} > 0$$

Indeed,

$$\sup_{q_h \in \widehat{Q}_h} \frac{\int_\Omega v_h \, \mathrm{div} q_h \, dx}{\|v_h\| \|q_h\|_{\mathrm{div}}} \geq \frac{\int_\Omega v_h \, \mathrm{div} q_h^v \, dx}{\|v_h\| \|q_h^v\|_{\mathrm{div}}} = \frac{\|v_h\|}{\|q_h\|_{\mathrm{div}}} \geq \frac{1}{\sqrt{1 + C^2}}.$$

Now, we refer to known results on the solvability of DMM, that can be summarized as follows: if the triangulations are "regular" and the discrete Inf-Sup condition holds, then the discrete formulation has a unique solution.

Projection type estimate for the dual problem

Since $\mathbf{p}$ is a maximizer, i.e.,

$$-\frac{1}{2} \| \mathbf{q} \|_*^2 \leq -\frac{1}{2} \| \mathbf{p} \|_*^2 \qquad \forall \mathbf{q} \in \mathbf{Q_f},$$

we find that

$$\int_\Omega \mathbf{A}^{-1}\mathbf{p} \cdot \mathbf{q}\, \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{q} \in \mathbf{Q_0},$$

where $\mathbf{Q_0}$ is the space of solenoidal functions. Therefore, for any $\mathbf{q} \in \mathbf{Q_f}$,

$$\frac{1}{2} \| \mathbf{q} - \mathbf{p} \|_*^2 = \frac{1}{2} \| \mathbf{q} \|_*^2 - \frac{1}{2} \| \mathbf{p} \|_*^2 + \int_\Omega \mathbf{A}^{-1}\mathbf{p} \cdot (\mathbf{p} - \mathbf{q}) \mathbf{dx} =$$

$$= \frac{1}{2} \| \mathbf{q} \|_*^2 - \frac{1}{2} \| \mathbf{p} \|_*^2 .$$

Let $\mathbf{Q_{fh}} = \mathbf{Q_f} \cap \widehat{\mathbf{Q}}_\mathbf{h}$. Note that $\mathbf{p_h} \in \mathbf{Q_{fh}}$ is also the maximizer of $-\frac{1}{2} \| \mathbf{q_{fh}} \|_*^2$ on $\mathbf{Q_{fh}}$, so that

$$\frac{1}{2} \| \mathbf{p_h} - \mathbf{p} \|_*^2 = \frac{1}{2} \| \mathbf{p_h} \|_*^2 - \frac{1}{2} \| \mathbf{p} \|_*^2 \leq \frac{1}{2} \| \mathbf{q_{fh}} \|_*^2 - \frac{1}{2} \| \mathbf{p} \|_*^2 =$$
$$= \frac{1}{2} \| \mathbf{q_{fh}} - \mathbf{p} \|_*^2 \quad \forall \mathbf{q_{fh}} \in \mathbf{Q_{fh}}.$$

Thus, we arrive at the first projection estimate

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \inf_{\mathbf{q_{fh}} \in \mathbf{Q_{fh}}} \| \mathbf{p} - \mathbf{q_{fh}} \|_* . \tag{148}$$

However, this projection error estimate has an obvious drawback. It is applicable only for a very narrow class of approximations: conforming (internal) approximations of the set $\mathbf{Q_f}$.

To obtain an estimate for a wider class, we first derive one auxiliary result.

A Modified DM problem

Take $\widetilde{\mathbf{f}} = \mathbf{div}(\widehat{\mathbf{q}}_\mathbf{h} - \mathbf{p})$ where $\widehat{\mathbf{q}}_h \in \widehat{\mathbf{Q}}_h$ and solve the modified DM problem

$$\int_\Omega \left(\mathbf{A}^{-1}\widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \cdot \widehat{\mathbf{q}}_\mathbf{h} + \widehat{\mathbf{u}}^\mathbf{f}_\mathbf{h}\mathbf{div}\widehat{\mathbf{q}}_\mathbf{h}\right)\mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_\mathbf{h} \in \widehat{\mathbf{Q}}_{\mathbf{0h}}, \tag{149}$$

$$\int_\Omega (\mathbf{div}\widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} + \widetilde{\mathbf{f}})\widehat{\mathbf{v}}_\mathbf{h}\mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}}_\mathbf{h} \in \widehat{\mathbf{V}}_\mathbf{h}. \tag{150}$$

Under the assumptions made $\widetilde{\mathbf{f}} \in \mathbf{P^0(T)}$, the above DM problem is solvable, and

$$\| \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \|^2_* + \int_\Omega \widehat{\mathbf{u}}^\mathbf{f}_\mathbf{h}\mathbf{div}\widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h}\mathbf{dx} = \mathbf{0},$$

$$\| \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \|^2_* \leq \|\widehat{\mathbf{u}}^\mathbf{f}_\mathbf{h}\| \|\mathbf{div}\widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h}\| = \|\widehat{\mathbf{u}}^\mathbf{f}_\mathbf{h}\| \|\widetilde{\mathbf{f}}\|$$

From here, we observe that

$$\bar{c}_1 \|\widehat{p}_h^f\|^2 \le \| \widehat{p}_h^f \|_*^2 \le \|\widehat{u}_h^f\| \|\widetilde{f}\|. \qquad (151)$$

By (146) and (147) we conclude that for $\widehat{u}_h^f$ we can find $\bar{q}_h$ in $\widehat{Q}_h$ such that

$$div \bar{q}_h + \widehat{u}_h^f = 0 \quad \text{and} \quad \|\bar{q}_h\| \le C \|\widehat{u}_h^f\|$$

Use $\bar{q}_h$ in the first identity (149). We have,

$$\int_\Omega \left(A^{-1} \widehat{p}_h^f \cdot \bar{q}_h + \widehat{u}_h^f div \bar{q}_h \right) dx = 0$$

Thus,

$$\|\widehat{u}_h^f\|^2 = \int_\Omega \widehat{u}_h^f div \bar{q}_h \le \| \widehat{p}_h^f \|_* \| \bar{q}_h \|_* \le$$

$$\le \bar{c}_2 \| \widehat{p}_h^f \|_* \|\bar{q}_h\| \le \bar{c}_2 C \| \widehat{p}_h^f \|_* \|\widehat{u}_h^f\|.$$

We observe that

$$\|\widehat{\mathbf{u}}_h^f\| \le \bar{c}_2 C \,|||\, \widehat{\mathbf{p}}_h^f \,|||_* \,. \tag{152}$$

Now, we use (151) and obtain

$$|||\, \widehat{\mathbf{p}}_h^f \,|||_*^2 \le \|\widehat{\mathbf{u}}_h^f\| \|\widetilde{\mathbf{f}}\| \le \bar{c}_2 C \,|||\, \widehat{\mathbf{p}}_h^f \,|||_* \, \|\widetilde{\mathbf{f}}\|.$$

so that

$$\bar{c}_1 \|\widehat{\mathbf{p}}_h^f\| \le |||\, \widehat{\mathbf{p}}_h^f \,|||_* \le \bar{c}_2 C \|\widetilde{\mathbf{f}}\|. \tag{153}$$

Hence,

$$\|\widehat{\mathbf{p}}_h^f\|_{\mathbf{div}}^2 = \|\widehat{\mathbf{p}}_h^f\|^2 + \|\mathbf{div}\widehat{\mathbf{p}}_h^f\|^2 \le (1 + \frac{c_2^2}{c_1^2} C^2) \|\widetilde{\mathbf{f}}\|^2. \tag{154}$$

We note that the estimates (152), (153), and (154) show that the modified DM problem is **stable**, i.e. its solutions $(\widehat{\mathbf{p}}_{\mathbf{h}}^{\mathbf{f}}, \widehat{\mathbf{u}}_{\mathbf{h}}^{\mathbf{f}})$ are bounded by the problem data uniformly with respect to $\mathbf{h}$.

If replace $\widetilde{\mathbf{f}}$ by $\mathbf{f}$, then we can derive the same stability estimate for the functions $(\widehat{\mathbf{p}}_{\mathbf{h}}, \widehat{\mathbf{u}}_{\mathbf{h}})$ that present an approximate solution of the original DM problem.

Projection estimates for fluxes

Now, we return to the projection error estimates. As we have seen

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \inf_{\mathbf{q_{fh}} \in \mathbf{Q_{fh}}} \|\mathbf{p} - \mathbf{q_{fh}}\|.$$

This estimate did not satisfy us because the set $\mathbf{Q_{fh}}$ is difficult to construct.
To avoid this drawback, we apply the following procedure.
Let $\boldsymbol{\eta}_h = \widehat{\mathbf{p}}^f_h + \widehat{\mathbf{q}}_h$, where $\widehat{\mathbf{q}}_h$ is *an arbitrary element of* $\widehat{\mathbf{Q}}_\mathbf{h}$.
We have,

$$\begin{aligned} \mathbf{div}\boldsymbol{\eta}_\mathbf{h} = \mathbf{div}\widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} + \mathbf{div}\widehat{\mathbf{q}}_\mathbf{h} = -\widetilde{\mathbf{f}} + \mathbf{div}\widehat{\mathbf{q}}_\mathbf{h} = \\ = \mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h}) + \mathbf{div}\widehat{\mathbf{q}}_\mathbf{h} = \mathbf{div}\mathbf{p} = -\mathbf{f}. \end{aligned}$$

Therefore, $\boldsymbol{\eta}_\mathbf{h} \in \mathbf{Q_f}$

Now, we recall the projection inequality and substitute in it $\boldsymbol{\eta}_h$:

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \| \mathbf{p} - \boldsymbol{\eta}_\mathbf{h} \|_* = \| \mathbf{p} - \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} - \widehat{\mathbf{q}}_\mathbf{h} \|_* \leq \tag{155}$$

$$\leq \| \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \|_* + \| \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \|_* \, . \tag{156}$$

Note that in the case considered $\mathbf{div}(\mathbf{p} - \mathbf{p_h}) = \mathbf{0}$, so that

$$\|\mathbf{p} - \mathbf{p_h}\|_{\mathbf{div}} = \|\mathbf{p} - \mathbf{p_h}\| \leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_\mathbf{1}} \| \mathbf{p} - \mathbf{p_h} \|_* \, .$$

Therefore,

$$\|\mathbf{p} - \mathbf{p_h}\|_{\mathbf{div}} \leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_\mathbf{1}} (\| \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \|_* + \| \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \|_*) \leq$$

$$(\mathbf{153}) \leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_\mathbf{1}} (\| \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \|_* + \bar{\mathbf{c}}_\mathbf{2} \mathbf{C} \|\widetilde{\mathbf{f}}\|).$$

But $\widetilde{\mathbf{f}} = \mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h})$.

Thus, we arrive at the estimate

$$\|\mathbf{p}-\mathbf{p_h}\|_{\mathbf{div}} \leq$$
$$\leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_{\mathbf{1}}}(\|\!|\, \mathbf{p}-\widehat{\mathbf{q}}_{\mathbf{h}} \,|\!\|_{*} + \bar{\mathbf{c}}_{\mathbf{2}}\mathbf{C}\|\mathbf{div}(\mathbf{p}-\widehat{\mathbf{q}}_{\mathbf{h}})\|) \quad \forall \widehat{\mathbf{q}}_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{h}}.$$

and, therefore,

$$\|\mathbf{p}-\mathbf{p_h}\|_{\mathbf{div}} \leq \bar{\mathbf{C}}_{\mathbf{p}} \inf_{\widehat{\mathbf{q}}_{\mathbf{h}}\in\widehat{\mathbf{Q}}_{\mathbf{h}}} \{\|\!|\, \mathbf{p}-\widehat{\mathbf{q}}_{\mathbf{h}} \,|\!\|_{*} + \|\mathbf{div}(\mathbf{p}-\widehat{\mathbf{q}}_{\mathbf{h}})\|\}. \quad (157)$$

where $\bar{\mathbf{C}}_{\mathbf{p}}$ depends on $\mathbf{C}, \bar{\mathbf{c}}_{\mathbf{1}}$, and $\bar{c}_2$ and does not depend on $\mathbf{h}$.

Projection type error estimates for $\widehat{\mathbf{u}} - \widehat{\mathbf{u}}_h$

We have

$$\int_{\Omega} \left(\mathbf{A}^{-1} \widehat{\mathbf{p}}_\mathbf{h} \cdot \widehat{\mathbf{q}}_\mathbf{h} + \widehat{\mathbf{u}}_\mathbf{h} \mathbf{div} \widehat{\mathbf{q}}_\mathbf{h} \right) \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_\mathbf{h} \in \widehat{\mathbf{Q}}_\mathbf{h}.$$

Since $\widehat{\mathbf{Q}}_\mathbf{h} \subset \mathbf{Q}$, we also have

$$\int_{\Omega} \left(\mathbf{A}^{-1} \mathbf{p} \cdot \widehat{\mathbf{q}}_\mathbf{h} + \mathbf{u} \mathbf{div} \widehat{\mathbf{q}}_\mathbf{h} \right) \mathbf{dx} = \mathbf{0}.$$

From here, we observe that

$$\int_{\Omega} \left(\mathbf{A}^{-1} (\widehat{\mathbf{p}}_\mathbf{h} - \mathbf{p}) \cdot \widehat{\mathbf{q}}_\mathbf{h} + (\widehat{\mathbf{u}}_\mathbf{h} - \mathbf{u}) \mathbf{div} \widehat{\mathbf{q}}_\mathbf{h} \right) \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_\mathbf{h} \in \widehat{\mathbf{Q}}_\mathbf{h}.$$

Denote

$$[\mathbf{u}]_{\mathbf{T}} = \frac{\mathbf{1}}{|\mathbf{T}|}\int_{\mathbf{T}} \mathbf{u}\mathbf{dx}, \qquad [\mathbf{u}]_{\mathbf{h}}(\mathbf{x}) = [\mathbf{u}]_{\mathbf{T_i}} \quad \text{if} \quad \mathbf{x} \in \mathbf{T_i}.$$

Since $\mathbf{div}\widehat{\mathbf{q}}_{\mathbf{h}}$ is constant on each $\mathbf{T_i}$, we rewrite the relation as follows:

$$\int_{\Omega} \left(\mathbf{A}^{-1}(\widehat{\mathbf{p}}_{\mathbf{h}} - \mathbf{p}) \cdot \widehat{\mathbf{q}}_{\mathbf{h}} + (\widehat{\mathbf{u}}_{\mathbf{h}} - [\mathbf{u}]_{\mathbf{h}})\mathbf{div}\widehat{\mathbf{q}}_{\mathbf{h}}\right)\mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{h}}.$$

Note that $[\mathbf{u}]_{\mathbf{h}} \in \widehat{\mathbf{V}}_{\mathbf{h}}$ and therefore $\bar{\mathbf{u}}_{\mathbf{h}} := \widehat{\mathbf{u}}_{\mathbf{h}} - [\mathbf{u}]_{\mathbf{h}} \in \widehat{\mathbf{V}}_{\mathbf{h}}$ Now, we exploit the compatibility and stability conditions (146) and (147) again. For $\bar{\mathbf{u}}_h$ one can find $\mathbf{q}'_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{h}}$ such that

$$\mathbf{div}\mathbf{q}'_{\mathbf{h}} + \bar{\mathbf{u}}_{\mathbf{h}} = \mathbf{0} \text{ and } \|\mathbf{q}'_{\mathbf{h}}\| \leq \mathbf{C}\|\bar{\mathbf{u}}_{\mathbf{h}}\|.$$

Let us use this function $\mathbf{q}'_{\mathbf{h}}$ in the integral relation. We have

$$\int_{\Omega} \left(\mathbf{A}^{-1}(\widehat{\mathbf{p}}_{\mathbf{h}} - \mathbf{p}) \cdot \mathbf{q}'_{\mathbf{h}} + \bar{\mathbf{u}}_{\mathbf{h}} \mathbf{div} \mathbf{q}'_{\mathbf{h}} \right) \mathbf{dx} = \mathbf{0}.$$

From here, we conclude that

$$\begin{aligned} \|\bar{\mathbf{u}}_{\mathbf{h}}\|^2 &= \left| \int_{\Omega} \mathbf{A}^{-1}(\widehat{\mathbf{p}}_{\mathbf{h}} - \mathbf{p}) \cdot \mathbf{q}'_{\mathbf{h}} \right| \leq \\ &\leq \| \widehat{\mathbf{p}}_{\mathbf{h}} - \mathbf{p} \|_* \| \mathbf{q}'_{\mathbf{h}} \|_* \leq \mathbf{C}\, \bar{\mathbf{c}}_2 \| \widehat{\mathbf{p}}_{\mathbf{h}} - \mathbf{p} \|_* \, \|\bar{\mathbf{u}}_{\mathbf{h}}\|. \end{aligned}$$

Thus,

$$\|\bar{\mathbf{u}}_{\mathbf{h}}\| = \|[\mathbf{u}]_{\mathbf{h}} - \widehat{\mathbf{u}}_{\mathbf{h}}\| \leq \mathbf{C}\, \bar{\mathbf{c}}_2 \| \widehat{\mathbf{p}}_{\mathbf{h}} - \mathbf{p} \|_* .$$

We have

$$\|\mathbf{u} - \widehat{\mathbf{u}}_\mathbf{h}\| \leq \|\mathbf{u} - [\mathbf{u}]_\mathbf{h}\| + \|[\mathbf{u}]_\mathbf{h} - \widehat{\mathbf{u}}_\mathbf{h}\| \leq$$
$$\leq \|\mathbf{u} - [\mathbf{u}]_\mathbf{h}\| + \mathbf{C}\,\bar{\mathbf{c}}_2 \,|\!|\!|\, \widehat{\mathbf{p}}_\mathbf{h} - \mathbf{p} \,|\!|\!|_*$$

Note that by the definition of $[\mathbf{u}]_\mathbf{h}$

$$\|\mathbf{u} - [\mathbf{u}]_\mathbf{h}\| \leq \|\mathbf{u} - \mathbf{v}_\mathbf{h}\| \qquad \forall \mathbf{v}_\mathbf{h} \in \widehat{\mathbf{V}}_\mathbf{h}.$$

From here, we observe that

$$\|\mathbf{u} - \widehat{\mathbf{u}}_\mathbf{h}\| \leq \mathbf{C}\,\bar{\mathbf{c}}_2 \,|\!|\!|\, \widehat{\mathbf{p}}_\mathbf{h} - \mathbf{p} \,|\!|\!|_* + \inf_{\mathbf{v}_\mathbf{h} \in \widehat{\mathbf{V}}_\mathbf{h}} \|\mathbf{u} - \mathbf{v}_\mathbf{h}\|$$

Recall (156) and observe that

$$|\!|\!|\, \mathbf{p} - \mathbf{p}_\mathbf{h} \,|\!|\!|_* \leq |\!|\!|\, \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \,|\!|\!|_* + |\!|\!|\, \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \,|\!|\!|_* \leq$$
$$|\!|\!|\, \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \,|\!|\!|_* + \bar{\mathbf{c}}_2\,\mathbf{C}\,\|\mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h})\|.$$

Then, we arrive at the projection type error estimate for the primal variable

$$\|\mathbf{u}-\widehat{\mathbf{u}}_\mathbf{h}\| \le \le \mathbf{C_u} \inf_{\widehat{\mathbf{q}}_\mathbf{h}\in\widehat{\mathbf{Q}}_\mathbf{h}} \Big\{ \|\!|\, \mathbf{p}-\widehat{\mathbf{q}}_\mathbf{h} \,|\!\|_* + \|\mathbf{div}(\mathbf{p}-\widehat{\mathbf{q}}_\mathbf{h})\| + + \inf_{\mathbf{v}_\mathbf{h}\in\widehat{\mathbf{V}}_\mathbf{h}} \|\mathbf{u}-\mathbf{v}_\mathbf{h}\| \Big\}, \quad (158)$$

where $\mathbf{C_u}$ depends on $\mathbf{C}, \bar{\mathbf{c}}_\mathbf{1}$, and $\bar{\mathbf{c}}_\mathbf{2}$ and does not depend on $\mathbf{h}$.
Estimates (157) and (158) lead to a qualified a priori convergence estimates provided that the solution possesses proper regularity.

Now, we proceed to the derivation of **functional type a posteriori error estimates** for the Primal Mixed and Dual Mixed methods.
Our analysis follows the lines of the paper
S. Repin and A. Smolianski, RJNAMM (2005).

Our aim is to find the difference between $(\mathbf{u}, \mathbf{p})$ and $(\mathbf{v}, \mathbf{q})$ in the respective energy norm, e.g. in the norm of the space $\mathbf{V} \times \mathbf{U}$.

A posteriori error estimates for PMM

A posteriori estimates for the mixed formulation are based on the relation that we have already derived:

$$\| \mathbf{p} - \widetilde{\mathbf{q}} \|_*^2 + \| \nabla(\mathbf{u} - \mathbf{v}) \|^2 = 2(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\widetilde{\mathbf{q}})) \,,$$

where $\widetilde{\mathbf{q}} \in Q_f$ and $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$.

Since the difference of the functionals in the right–hand side can be estimated by the known way, we arrive at the estimate

$$\| \mathbf{p} - \widetilde{\mathbf{q}} \|_*^2 + \| \nabla(\mathbf{u} - \mathbf{v}) \|^2 \leq \mathbf{2}(\mathbf{1} + \beta)\mathbf{D}(\nabla \mathbf{v}, \mathbf{y}) + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \mathbf{C}^2 \left(\|\mathbf{div y} + \mathbf{f}\|^2 + \|\mathbf{y} \cdot \mathbf{n} - \mathbf{F}\|_{\partial_2 \Omega}^2 \right) , \quad (159)$$

where $\mathbf{y} \in \widehat{\mathbf{Q}}^+$, $\widetilde{\mathbf{q}} \in \mathbf{Q_f}$ and $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$ are arbitrary functions and β is any positive number.

Thus, for the error in the primal variable we have

$$\| \nabla(\mathbf{u} - \mathbf{u_h}) \|^2 \leq \mathbf{2}(\mathbf{1} + \beta)\mathbf{D}(\nabla \mathbf{u_h}, \mathbf{y}) + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \mathbf{C}^2 \left(\|\mathbf{div y} + \mathbf{f}\|^2 + \|\mathbf{y} \cdot \mathbf{n} - \mathbf{F}\|^2_{\partial_2 \Omega} \right) . \quad (160)$$

where $\mathbf{C}$ is a constant in the inequality

$$\|\mathbf{w}\|^2 + \|\mathbf{w}\|^2_{\partial_2 \Omega} \leq \mathbf{C}^2 \| \nabla \mathbf{w} \|^2 \quad \forall \mathbf{w} \in \mathbf{V_0} .$$

A posteriori estimate for the dual variable

Error estimates for the dual variable in the dual energy norm $\|\!| \cdot |\!|_*$ can be obtained by the arguments similar to those used above.
Let $\mathbf{q} \in \mathbf{Y}^*$ be an approximation of $\mathbf{p}$. For any $\widetilde{\mathbf{q}} \in \mathbf{Q}_\mathbf{f}^*$, we obtain (from the triangle inequality and Young inequalities with $\gamma > 0$)

$$\|\!| \mathbf{q} - \mathbf{p} |\!|_*^2 \leq (1+\gamma) \|\!| \mathbf{q} - \widetilde{\mathbf{q}} |\!|_*^2 + \left(1 + \frac{1}{\gamma}\right) \|\!| \widetilde{\mathbf{q}} - \mathbf{p} |\!|_*^2 .$$

Use the fact that $\|\!| \widetilde{\mathbf{q}} - \mathbf{p} |\!|_*^2 \leq 2\,(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\widetilde{\mathbf{q}}))$.

Therefore, for any $\mathbf{w} \in \mathbf{V_0} + \mathbf{u_0}$, we have

$$\| \mathbf{q} - \mathbf{p} \|_*^2 \leq (1+\gamma) \| \mathbf{q} - \widetilde{\mathbf{q}} \|_*^2 + 2\left(1 + \frac{1}{\gamma}\right)(\mathbf{J}(\mathbf{u}) - \mathbf{I}^*(\widetilde{\mathbf{q}})) \leq$$

$$\leq (1+\gamma) \| \mathbf{q} - \widetilde{\mathbf{q}} \|_*^2 + 2\left(1 + \frac{1}{\gamma}\right)(\mathbf{J}(\mathbf{w}) - \mathbf{I}^*(\widetilde{\mathbf{q}})) =$$

Recall that

$$\mathbf{J}(\mathbf{w}) - \mathbf{I}^*(\widetilde{\mathbf{q}}) \leq (1+\beta)\mathbf{D}(\nabla \mathbf{w}, \mathbf{y}) + \left(1 + \frac{1}{\beta}\right)\frac{1}{2} \| \mathbf{q} - \widetilde{\mathbf{q}} \|_*^2$$

so that the right–hand side is estimated by

$$(1+\gamma)\left(1 + \frac{1}{\gamma} + \frac{1}{\beta\gamma}\right) \mathbf{d_{Q_f}}^2(\mathbf{q}) + 2(1+\beta)\left(1 + \frac{1}{\gamma}\right)\mathbf{D}(\nabla \mathbf{w}, \mathbf{q}).$$

If $\mathbf{q} \in \mathbf{U}$, then $\mathbf{d_{Q_f}}$ is given by the negative norm and we obtain

$$\begin{aligned} \| \mathbf{q} - \mathbf{p} \|_*^2 \quad &\leq (1+\gamma)\left(1 + \frac{1}{\gamma} + \frac{1}{\beta\gamma}\right) \| \mathbf{f} + \mathbf{div q} \|_{\Omega}^2 + \\ &+ (1+\beta)\left(1 + \frac{1}{\gamma}\right) \mathbf{D}(\nabla \mathbf{w}, \mathbf{q}). \end{aligned} \tag{161}$$

If $\mathbf{q} \in \mathbf{Q}$ then we have (for the Dirichlét problem)

$$\begin{aligned} \| \mathbf{q} - \mathbf{p} \|_*^2 \quad &\leq (1+\gamma)\left(1 + \frac{1}{\gamma} + \frac{1}{\beta\gamma}\right) \mathbf{C_\Omega} \| \mathbf{f} + \mathbf{div q} \| + \\ &+ (1+\beta)\left(1 + \frac{1}{\gamma}\right) \mathbf{D}(\nabla \mathbf{w}, \mathbf{q}). \end{aligned} \tag{162}$$

This estimate holds for any positive parameters β, γ, and any $\mathbf{w} \in \mathbf{V_0} + \mathbf{u_0}$. Here $\mathbf{w}$ is a "free" function in $\mathbf{V_0} + \mathbf{u_0}$. This "freedom" can be used to make the estimate sharper.

Application of (161) to DM approximations leads (after optimization with respect to scalar parameters) to the estimate

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \sqrt{2}\mathbf{D}^{1/2}(\nabla \mathbf{w}, \mathbf{y}) + \| \mathbf{y} - \mathbf{p_h} \|_* + \mathbf{2C}\left(\|\mathbf{div y} + \mathbf{f}\|^2 + \|\mathbf{y}\cdot\mathbf{n} - \mathbf{F}\|^2 \right)^{1/2} . \quad (163)$$

Here $\mathbf{w}$ is an arbitrary function from $\mathbf{V_0} + \mathbf{u_0}$ and $\mathbf{y}$ is an arbitrary function from $\widehat{\mathbf{Q}}^+$. If $\mathbf{y} = \mathbf{A}\nabla\mathbf{u}$ and $w = u$, then the right-hand side of (163) coincides with the left-hand side, i.e. is exact in the sense that there exist such "free variables" that the inequality holds as the equality.

A directly computable upper bound of $\| \mathbf{p} - \mathbf{p_h} \|_*$ is given by (163), if we set

$$\mathbf{v} = \mathbf{u_h}, \quad \text{and} \quad \mathbf{y} = \mathcal{G}_\mathbf{h}\mathbf{p_h},$$

where $\mathcal{G}_\mathbf{h} : \mathbf{Q_h} \to \widehat{\mathbf{Q}}^+$ is a certain projection operator (some examples such operators has been already discussed in the previous lectures). We have

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \sqrt{2}\mathbf{D}^{1/2}(\nabla \mathbf{u_h}, \mathcal{G}_\mathbf{h}\mathbf{p_h}) + \| \mathcal{G}_\mathbf{h}\mathbf{p_h} - \mathbf{p_h} \|_* + 2\mathbf{C} \left(\|\mathbf{div}\mathcal{G}_\mathbf{h}\mathbf{p_h} + \mathbf{f}\|^2 + \|\mathcal{G}_\mathbf{h}\mathbf{p_h} \cdot \mathbf{n} - \mathbf{F}\|^2 \right)^{1/2}.$$

Projection from $\mathbf{Q_h}$ onto $\widehat{\mathbf{Q}}^+$

If $\mathbf{p_h}$ is a piecewise-constant vector field on a simplicial mesh $\mathcal{T}_h$, then, Raviart–Thomas elements (e.g., $\mathbf{RT^0}$–elements) can be used in order to define the mapping $\mathcal{G}$.
Assume that the Ω has a polygonal boundary, and the latter is exactly matched by the triangulation $\mathcal{T}_h$. Let $\mathbf{T_i}$ and $\mathbf{T_j}$ be two neighboring simplexes with the common edge $\mathbf{E_{ij}}$. Let $\mathbf{q_h}$ be a piecewise constant vector-valued function that has the values $\mathbf{q_i}$ and $\mathbf{q_j}$ on $\mathbf{T_i}$ and $\mathbf{T_j}$ respectively. Let $\mathbf{E_{ij}}$ be the common edhge with the unit normal $\mathbf{n_{ij}}$ oriented from $\mathbf{T_i}$ to $\mathbf{T_j}$ if $\mathbf{i} > \mathbf{j}$.

How to define the common value $\widetilde{\mathbf{q}}_{ij} \cdot \mathbf{n_{ij}}$ on $\mathbf{E_{ij}}$?

One possible option is as follows:

$$\widetilde{\mathbf{q}}_{\mathbf{ij}} \cdot \mathbf{n}_{\mathbf{ij}} = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{q}_{\mathbf{i}} + \mathbf{q}_{\mathbf{j}}) \cdot \mathbf{n}_{\mathbf{ij}} \,,$$

Another option is

$$\widetilde{\mathbf{q}}_{\mathbf{ij}} \cdot \mathbf{n}_{\mathbf{ij}} = \frac{|\mathbf{T}_{\mathbf{i}}|\mathbf{q}_{\mathbf{i}} + |\mathbf{T}_{\mathbf{j}}|\mathbf{q}_{\mathbf{j}}}{|\mathbf{T}_{\mathbf{i}}| + |\mathbf{T}_{\mathbf{j}}|} \cdot \mathbf{n}_{\mathbf{ij}} \,,$$

where $|\mathbf{T}_{\mathbf{i}}|$ and $|\mathbf{T}_{\mathbf{j}}|$ are the areas of $\mathbf{T}_{\mathbf{i}}$ and $\mathbf{T}_{\mathbf{j}}$. We repeat this procedure for all internal edges of $\mathcal{T}_{\mathbf{h}}$.
If $\mathbf{E}_{\mathbf{i0}} \in \partial_{\mathbf{1}}\mathbf{\Omega}$, then we set $\widetilde{\mathbf{q}}_{i0} \cdot \mathbf{n}_{\mathbf{i0}} = \mathbf{q}_{\mathbf{i0}} \cdot \mathbf{n}_{\mathbf{i0}}$. If $\mathbf{E}_{\mathbf{i0}} \in \partial_{\mathbf{2}}\mathbf{\Omega}$, then

$$\widetilde{\mathbf{q}}_{\mathbf{i0}} \cdot \mathbf{n}_{\mathbf{i0}} = \frac{\mathbf{1}}{|\mathbf{E}_{\mathbf{i0}}|} \int_{\mathbf{E}_{\mathbf{i0}}} \mathbf{F}\, \mathbf{ds} \,.$$

Here $|\mathbf{E}_{\mathbf{i0}}|$ is the length of the edge $\mathbf{E}_{\mathbf{i0}}$.

Thus, all the normal components $\widetilde{\mathbf{q}}_{ij} \cdot \mathbf{n_{ij}}$ on internal and external edges are defined. By prolongation inside all $\mathbf{T_i}$, with the help of $\mathbf{RT}_0$-approximations we obtain the function a piecewise affine function, which has continuous normal components at all the edges and piecewise constant normal components on $\partial\Omega$.
Therefore, we, in fact, have constructed a mapping $\mathbf{q_h} \longrightarrow \widetilde{\mathbf{q}}_\mathbf{h}$ such that

$$\widetilde{\mathbf{q}}_\mathbf{h} = \mathcal{G}_\mathbf{h}\mathbf{q_h} \in \widehat{\mathbf{Q}}^+ .$$

A posteriori estimates for DMM

An a posteriori estimate for the flux $\widehat{\mathbf{p}}_\mathbf{h}$ readily follows from the general estimate

$$\begin{aligned} \tfrac{1}{2} \| \mathbf{y} - \mathbf{p} \|_*^2 \quad &\le (1+\gamma) \left(1 + \frac{1}{\gamma} + \frac{1}{\beta\gamma} \right) \| \mathbf{f} + \mathbf{div}\, \mathbf{y} \|^2 + \\ &+ (1+\beta) \left(1 + \frac{1}{\gamma} \right) \mathbf{D}(\nabla \mathbf{w}, \mathbf{y}). \end{aligned}$$

We set $\mathbf{y} = \widehat{\mathbf{p}}_\mathbf{h} \in \widehat{\mathbf{Q}}^+$. Since $\widehat{\mathbf{p}}_h$ is a piecewise polynomial function, it has a summable trace on $\partial_2\Omega$. Then, we estimate $\| \mathbf{f} + \mathbf{div}\, \mathbf{y} \|$ from above in the same way as before.

Minimization with respect to γ and $\boldsymbol{\beta}$ leads to the estimate

$$\| \mathbf{p} - \widehat{\mathbf{p}}_\mathbf{h} \|_* \leq \sqrt{2}\mathbf{D}^{1/2}(\nabla \mathbf{w}, \widehat{\mathbf{p}}_\mathbf{h}) + \\ + 2\mathbf{C}\left(\|\mathbf{div}\widehat{\mathbf{p}}_\mathbf{h} + \mathbf{f}\|^2 + \|\widehat{\mathbf{p}}_\mathbf{h} \cdot \mathbf{n} - \mathbf{F}\|^2_{\partial_2\Omega} \right)^{1/2}, \tag{164}$$

where $\mathbf{w}$ is an arbitrary function from $\mathbf{V_0} + \mathbf{u_0}$.

Assume that Ω is a polygonal domain decomposed into a regular collection of simplexes. If $\widehat{\mathbf{p}}_h$ is constructed by means of $\mathbf{RT}_0$-elements, then

$$\int_\Omega (\mathbf{div}\widehat{\mathbf{p}}_h + \mathbf{f})\mathbf{w}_h \, \mathbf{dx} = 0 \quad \forall \mathbf{w}_h \in \widehat{\mathbf{V}}_h \subset \widehat{\mathbf{V}} \,, \tag{165}$$

where the subspace $\widehat{\mathbf{V}}_h$ contains piecewise constant functions.
Therefore, on each element $\mathbf{T_i}$

$$\mathbf{div}\widehat{\mathbf{p}}_h = -\frac{1}{|\mathbf{T_i}|}\int_{\mathbf{T_i}} \mathbf{f} \, \mathbf{dx} \,. \tag{166}$$

Let us define by $[\mathbf{f}]$ the function whose values on $\mathbf{T_i}$ coincide with the mean values of $\mathbf{f}$ on $\mathbf{T_i}$. It is clear that $[\mathbf{f}] \in \widehat{\mathbf{V}}_\mathbf{h}$ Then, we have

$$\mathbf{div}\widehat{\mathbf{p}}_\mathbf{h} = -[\mathbf{f}] \quad \text{on every} \quad \mathbf{T_i}.$$

Estimate (165) is valid for any approximate flux $\widehat{\mathbf{p}}_\mathbf{h}$ from $\widehat{\mathbf{Q}}^+$. If $\widehat{\mathbf{p}}_\mathbf{h}$ belongs to the narrower set $\widehat{\mathbf{Q}}_\mathbf{F}$ (as, e.g., it would be in the discrete dual mixed method if $\mathbf{f} = [\mathbf{f}]$ and pure Dirichlét conditions) then the last norm in (165) would be identically zero.
It cannot, however, be expected, when $\widehat{\mathbf{p}}_\mathbf{h}$ is constructed in the space $\mathbf{RT_0}$, unless the function $\mathbf{F}$ is a constant on $\partial_\mathbf{2}\mathbf{\Omega}$.

Remark. The problem of taking into account the essential boundary condition for the flux variable

$$\widehat{\mathbf{p}} \cdot \mathbf{n} = \mathbf{F} \quad \text{on} \quad \partial_2 \Omega$$

in the dual mixed method is not at all easy and, usually, leads to a non-conforming approximation $\widehat{\mathbf{p}}_\mathbf{h}$
(see, e.g., I. Babuska and G. N. Gatica,
On the mixed finite element method with Lagrange multipliers.
Numer. Meth. PDE, 2003).
However, (165) still works for such (nonconforming) approximations of the flux !

One simple nonconforming version of the discrete dual method, particularly suited for the lowest-order Raviart-Thomas approximation is as follows. Instead of requiring $\widehat{\mathbf{p}}_\mathbf{h} \in \widehat{\mathbf{Q}}_\mathbf{F}$, we impose a weaker condition

$$\widehat{\mathbf{p}}_\mathbf{h} \cdot \mathbf{n}\big|_{\mathbf{E_{i0}}} = \frac{\mathbf{1}}{|\mathbf{E_{i0}}|} \int_{\mathbf{E_{i0}}} \mathbf{F}\, \mathbf{ds} \tag{167}$$

on every edge $\mathbf{E_{i0}} \in \partial_\mathbf{2}\mathbf{\Omega}$.
If now we denote by $[\mathbf{F}]$ the piecewise constant function defined on the set of edges forming $\partial_\mathbf{2}\mathbf{\Omega}$ and whose value on every edge $\mathbf{E_{i0}} \in \partial_\mathbf{2}\mathbf{\Omega}$ is equal to the mean value of $\mathbf{F}$ on that edge, we can write that $\widehat{\mathbf{p}}_\mathbf{h} \cdot \mathbf{n} = [\mathbf{F}]$ for all $\mathbf{E_{i0}} \in \partial_\mathbf{2}\mathbf{\Omega}$.

In this case, **nonconformity errors will be automatically accounted in the functional a posteriori estimate** (165). Indeed, we obtain

$$\| \mathbf{p} - \widehat{\mathbf{p}}_\mathbf{h} \|_* \leq \sqrt{2}\mathbf{D}^{1/2}(\nabla \mathbf{v}, \widehat{\mathbf{p}}_\mathbf{h}) + \mathbf{2C}\left(\|\mathbf{f} - [\mathbf{f}]\|^2 + \|\mathbf{F} - [\mathbf{F}]\|^2_{\partial_2 \Omega} \right)^{1/2} . \quad (168)$$

How to choose in (168) the function $\mathbf{w} \in \mathbf{V_0} + \mathbf{u_0}$.

The simplest way is to use the function $\widehat{\mathbf{u}}_\mathbf{h} \in \widehat{\mathbf{V}}_\mathbf{h}$ available from the solution of the discrete dual mixed problem and to construct a suitable projection operator $\mathcal{P}_\mathbf{h} : \widehat{\mathbf{V}}_\mathbf{h} \to \mathbf{V_0} + \mathbf{u_0}$. Again, the projection can be easily accomplished with a simple averaging.

Projection from $\widehat{\mathbf{V}}_\mathbf{h}$ onto $\mathbf{V_0} + \mathbf{u_0}$.

In order to find $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$, it is sufficient to find $\mathbf{w} \in \mathbf{V_0}$ in the representation $\mathbf{v} = \mathbf{w} + \mathbf{u_0}$ (the function $\mathbf{u_0}$ is given). Using the computed piecewise-constant function $\widehat{\mathbf{u}}_\mathbf{h}$, we define $\mathbf{w}_\mathbf{h} \in \mathbf{V_0}$ as follows.

We set

$$\mathbf{w_h(x_k)} = \frac{\sum\limits_{\mathbf{s=1}}^{\mathbf{N_k}} |\mathbf{T_s^{(k)}}| \cdot \widehat{\mathbf{u}}_\mathbf{h}\big|_{\mathbf{T_s^{(k)}}}}{\sum\limits_{\mathbf{s=1}}^{\mathbf{N_k}} |\mathbf{T_s^{(k)}}|} - \mathbf{u_0(x_k)} \qquad (169)$$

for any internal node $\mathbf{x_k}$ and when $\mathbf{x_k} \in \partial_\mathbf{2}\mathbf{\Omega}$. Here $\mathbf{T_s^{(k)}}$, $\mathbf{s} = \overline{\mathbf{1}, \mathbf{N_k}}$, are the elements containing the vertex $\mathbf{x_k}$, and we have assumed that the function $\mathbf{u_0}$ has a sufficient regularity, so that its point values are defined.

If the node $\mathbf{x_k} \in \partial_\mathbf{1}\mathbf{\Omega}$, we simply set $\mathbf{w_h(x_k)} = \mathbf{0}$.

Thus, using the nodal values of $\mathbf{w_h}$ and the piecewise-linear continuous finite element approximation on the mesh $\mathcal{T}_\mathbf{h}$ we define the function

$$\mathbf{w_h} + \mathbf{u_0} = \mathcal{P}_\mathbf{h}\widehat{\mathbf{u}}_\mathbf{h} \in \mathbf{V_0} + \mathbf{u_0}\,.$$

Hence, from (168) one obtains

$$\bar{\mathbf{c}}_1 \|\mathbf{p} - \widehat{\mathbf{p}}_\mathbf{h}\| \leq \| \mathbf{p} - \widehat{\mathbf{p}}_\mathbf{h} \|_* \leq$$

$$\sqrt{2}\mathbf{D}^{1/2}(\nabla(\mathcal{P}_\mathbf{h}\widehat{\mathbf{u}}_\mathbf{h}), \widehat{\mathbf{p}}_\mathbf{h}) + \mathbf{2C}\left(\|\mathbf{f} - [\mathbf{f}]\|^2 + \|\mathbf{F} - [\mathbf{F}]\|^2_{\partial_2\Omega}\right)^{1/2}, \quad (170)$$

which, together with the obvious relation

$$\|\mathbf{div}(\widehat{\mathbf{p}} - \widehat{\mathbf{p}}_\mathbf{h})\| = \| - \mathbf{f} - \mathbf{div}\widehat{\mathbf{p}}_\mathbf{h}\| = \|\mathbf{f} - [\mathbf{f}]\|$$

leads to the upper bound for $\|\widehat{\mathbf{p}} - \widehat{\mathbf{p}}_\mathbf{h}\|_\mathbf{div}$:

Theorem

Let $(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) \in \widehat{\mathbf{V}} \times \widehat{\mathbf{Q}}_{\mathbf{F}}$ *be the exact solution of the dual mixed problem and* $(\widehat{\mathbf{u}}_{\mathbf{h}}, \widehat{\mathbf{p}}_{\mathbf{h}}) \in \widehat{\mathbf{V}}_{\mathbf{h}} \times \widehat{\mathbf{Q}}_{\mathbf{Fh}}$ *the solution of the discrete dual mixed problem with* $\widehat{\mathbf{Q}}_{\mathbf{Fh}}$ *being the Raviart-Thomas space* $\mathbf{RT}^{\mathbf{0}}$.
Then, the following estimate holds true:

$$\|\widehat{\mathbf{p}} - \widehat{\mathbf{p}}_{\mathbf{h}}\|_{\mathbf{div}} \leq \| \mathbf{A}\nabla(\mathcal{P}_{\mathbf{h}}\widehat{\mathbf{u}}_{\mathbf{h}}) - \widehat{\mathbf{p}}_{\mathbf{h}} \|_{*} + (\mathbf{2C}+\mathbf{1})\|\mathbf{f} - [\mathbf{f}]\| + \mathbf{2C}\|\mathbf{F} - [\mathbf{F}]\|_{\partial_2\Omega}\,, \quad (171)$$

where $\mathcal{P}_{\mathbf{h}} : \widehat{\mathbf{V}}_{\mathbf{h}} \to \mathbf{V_0} + \mathbf{u_0}$ *is the projection (averaging) operator introduced above and* $[\mathbf{f}]$ *and* $[\mathbf{F}]$ *are the averaged functions.*

Remark. The first and the second terms in (171), being computed elementwise, can serve as local error indicators.

A sharper estimate can be obtained by the minimization of the Majorant with respect to $\mathbf{v}$. Here, we can restrict ourselves to certain subspace V_h, i.e.,

$$\|\widehat{\mathbf{p}} - \widehat{\mathbf{p}}_\mathbf{h}\|_{\mathbf{div}} \leq \inf_{\mathbf{v_h} \in \mathbf{V_h}} \| \mathbf{A}\nabla(\mathbf{v_h}) - \widehat{\mathbf{p}}_\mathbf{h} \|_* + (\mathbf{2C}+\mathbf{1})\|\mathbf{f} - [\mathbf{f}]\| + \mathbf{2C}\|\mathbf{F} - [\mathbf{F}]\|_{\partial_2\Omega} . \quad (172)$$

By (160) we can also the squared norm of the error of the averaged solution $\mathcal{P}_\mathbf{h}\widehat{\mathbf{u}}_\mathbf{h}$ using the computed flux approximation $\widehat{\mathbf{p}}_\mathbf{h}$:

$$\| \nabla(\mathbf{u} - \mathcal{P}_\mathbf{h}\widehat{\mathbf{u}}_\mathbf{h}) \|^2 \leq \mathbf{2}(\mathbf{1}+\beta)\mathbf{D}(\nabla(\mathcal{P}_\mathbf{h}\widehat{\mathbf{u}}_\mathbf{h}), \widehat{\mathbf{p}}_\mathbf{h}) + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right)\mathbf{C}^2(\|\mathbf{f} - [\mathbf{f}]\|^2 + \|\mathbf{F} - [\mathbf{F}]\|^2_{\partial_2\Omega}), \quad (173)$$

where $\beta > \mathbf{0}$ is an arbitrary number that can be used to minimize the right-hand side of (173) and to obtain the estimate for the norm of the error.

A sharper estimate may be obtained, if one spends some time on the minimization of the right-hand side of (173) with respect to the dual variable $\mathbf{y}$ over some finite-dimensional subspace of $\widehat{\mathbf{Q}}^+$.

Remark.
If one has the solutions of both the primal and the dual mixed problems, the flux approximation $\widehat{\mathbf{p}}_\mathbf{h}$ can be substituted into (160) to immediately yield the error estimate for the primal variable (which is the most important in the primal mixed method), while the approximation $\mathbf{u}_\mathbf{h}$ can be used in (171) to bring the error estimate for the dual variable (which is the most important in the dual mixed method).

Lecture 7

MIXED FEM ON DISTORTED MESHES

The Plan

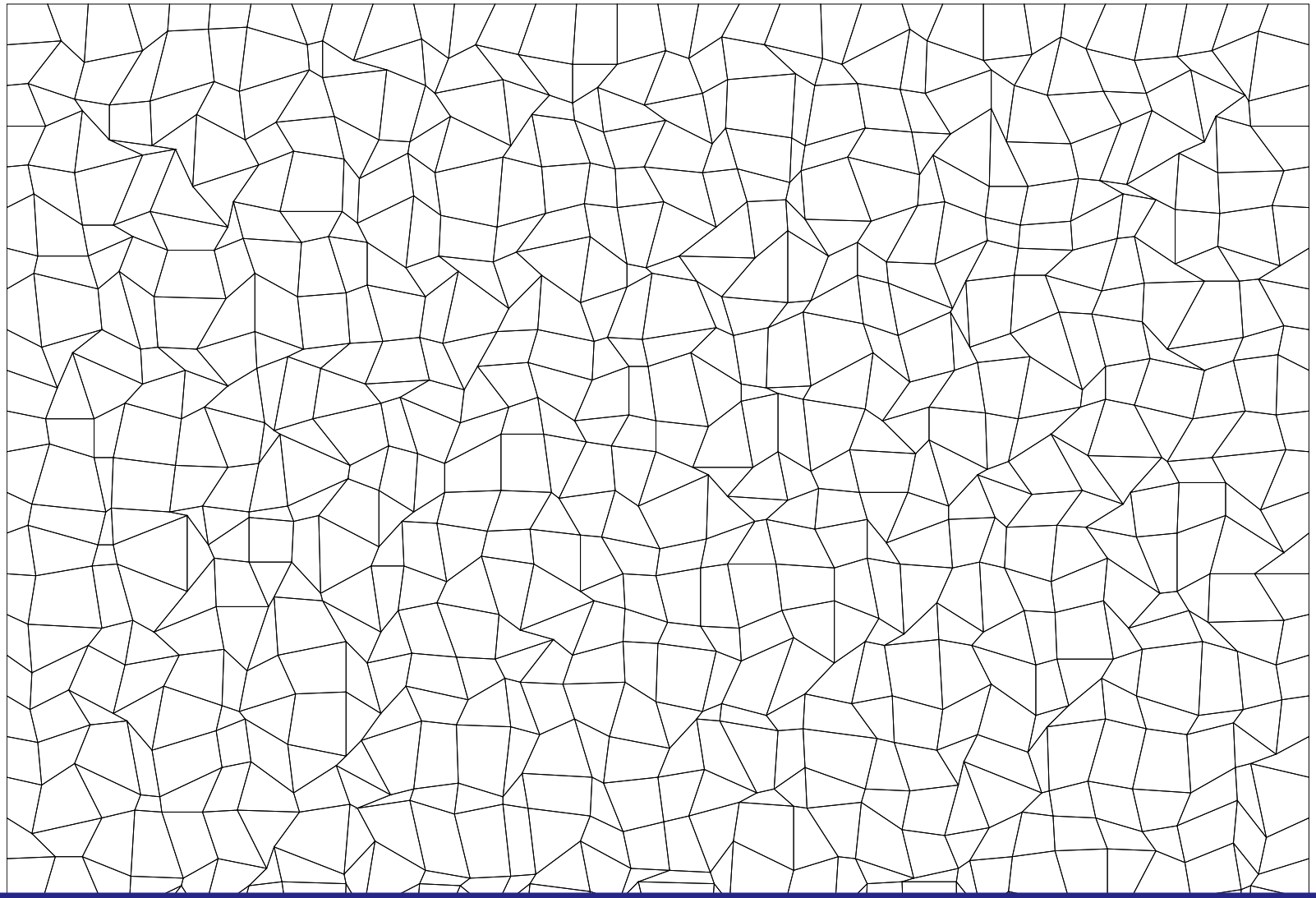
1. Approximations on distorted meshes
2. Inf-sup condition
3. A priori rate convergence estimates
4. A posteriori estimates

Distorted meshes

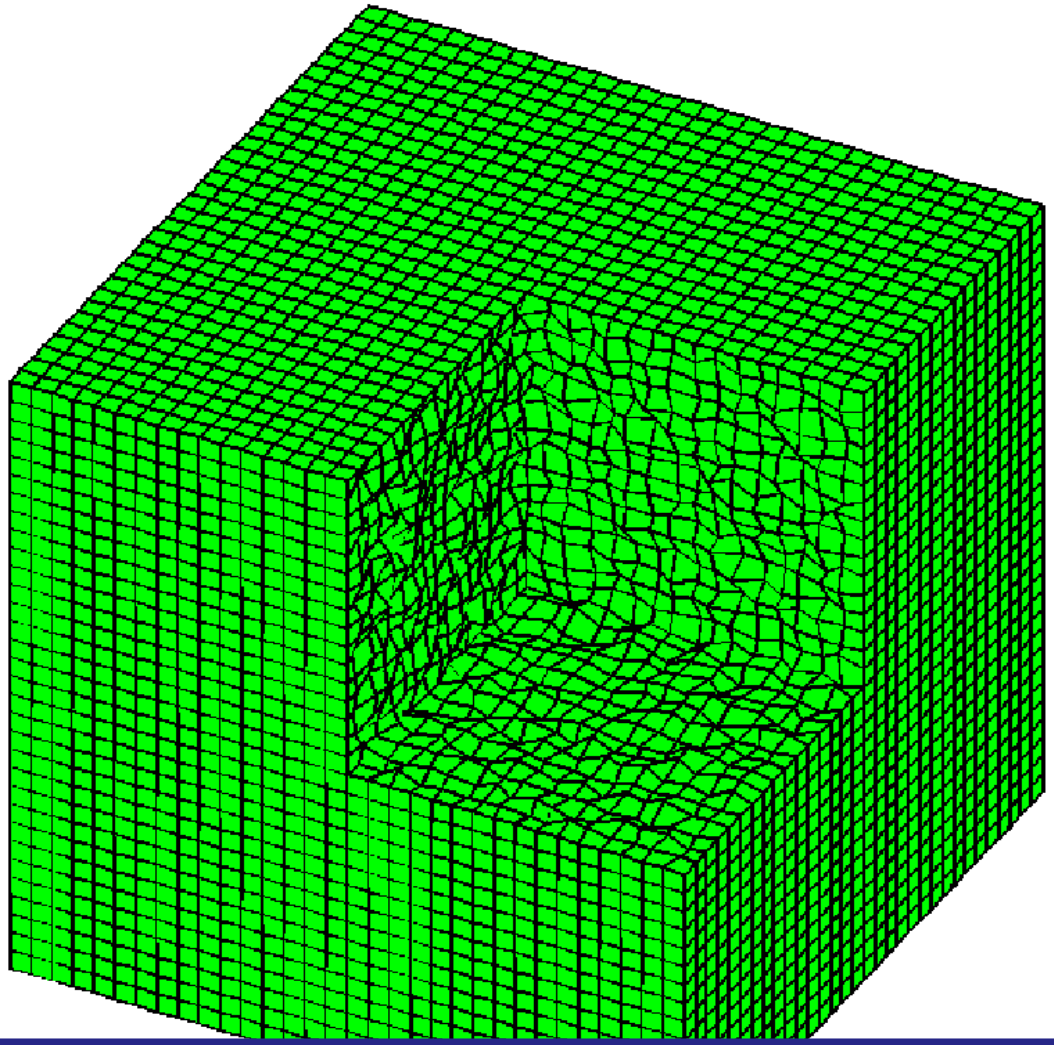
Distorted meshes may arise by several ways:

- Forcibly, due to the form of the given data;
- In the process of mesh refinement;
- In the process of adaptation to physical structure (e.g. layers of different materials)

Distorted mesh in 2D



Distorted mesh in 3D



How to work with such meshes?

Straightforward way:

Construct a regular mesh composed of standard elements that contains the given distorted one and solve the problem:

Drawback:

The problem obtained may have a very large dimension

We consider another *modus operandi* that is based upon a certain **aggregation procedure and reduction of some degrees of freedom.**

Basic problem

Principal ideas of the approach we discuss on the paradigm of the classical problem

$$-\mathbf{\Delta u} = \mathbf{f}, \qquad \text{in } \Omega \tag{174}$$

$$\frac{\partial \mathbf{u}}{\partial \nu} = \mathbf{0} \qquad \text{on } \partial\mathbf{\Omega}, \tag{175}$$

Ω is a bounded connected polygonal (polyhedral) domain in $\mathbb{R}^d$ ($d = 2, 3$), where ν is the outward unit normal vector to the boundary $\partial\Omega$ and it is assumed that

$$\mathbf{f} \in \mathbf{L_2(\Omega)}, \quad \int_{\mathbf{\Omega}} \mathbf{f}\, \mathbf{dx} = \mathbf{0}. \tag{176}$$

However, the approach can be extended to problems with the elliptic operator $\mathbf{div A}\nabla$ and other types of boundary conditions.

To introduce a minimax formulation of the problem (174)–(175), we define the space of square intergable functions with zero mean

$$\mathbf{V}(\Omega) := \left\{ \mathbf{v} \in \mathbf{L}_2(\Omega) \;||\; \int_\Omega \mathbf{v} d\mathbf{x} = \mathbf{0} \right\},$$

for the primal variable u and the set

$$\mathbf{Q}(\Omega) := \{ \mathbf{q} \in \mathbf{H}(\Omega, \mathrm{div}) \;||\; \mathbf{q} \cdot \nu = \mathbf{0} \text{ on } \partial\Omega \},$$

for the dual variable p, where the boundary condition is understood in a generalized sense and $H(\Omega, \mathrm{div})$ is the Hilbert space of vector–valued functions with the scalar product and the norm defined by the relations

$$(\mathbf{p}, \mathbf{q}) := \int_\Omega (\mathbf{p} \cdot \mathbf{q} + \mathrm{div}\mathbf{p}\, \mathrm{div}\mathbf{q})\, \mathbf{dx}, \qquad \|\mathbf{p}\|_{\mathrm{div},\Omega} := \sqrt{(\mathbf{p}, \mathbf{p})}.$$

On $\mathbf{V} \times \mathbf{Q}$, we define the Lagrangian

$$\mathbf{L}(\mathbf{v}, \mathbf{q}) := -\int_{\Omega} \left(\mathbf{v}\mathrm{div}\mathbf{q} + \frac{\mathbf{1}}{\mathbf{2}}|\mathbf{q}|^2 + \mathbf{f}\mathbf{v} \right) \mathbf{dx}.$$

As we have seen, the saddle–point problem: find $(\mathbf{u}, \mathbf{p}) \in \mathbf{V}(\Omega) \times \mathbf{Q}(\Omega)$ such that

$$\mathbf{L}(\mathbf{u}, \mathbf{q}) \leq \mathbf{L}(\mathbf{u}, \mathbf{p}) \leq \mathbf{L}(\mathbf{v}, \mathbf{p}) \quad \forall \mathbf{v} \in \mathbf{V},\ \mathbf{q} \in \mathbf{Q}$$

has a unique solution, which satisfies the equations

$$\int_{\Omega} (\mathrm{div}\mathbf{p} + \mathbf{f})\mathbf{w}\,\mathbf{dx} = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V}(\Omega), \tag{177}$$

$$\int_{\Omega} (\mathbf{p} \cdot \mathbf{q} + \mathbf{u}\,\mathrm{div}\mathbf{q})\,\mathbf{dx} = \mathbf{0} \quad \forall \mathbf{q} \in \mathbf{Q}(\Omega). \tag{178}$$

Discrete problems

Let $\mathfrak{T}_h$ be a partitioning of Ω into polygonal (polyhedral) cells n_s, $s = 1, 2, \ldots, N$, $h \longrightarrow 0$ as $N \to +\infty$.
We assume that
(a) partitioning is conforming in the sense that interfaces $\mathbf{E_{st}}$ between cells n_s and n_t are straight segments for $d = 2$ and simply connected polygons for $\mathbf{d = 3}$
(b) partitioning is quasiuniform and the cells are regularly shaped. The first assumption means that there exist two positive numbers α_1 and α_2 such that

$$\alpha_1 \mathbf{N}^{-\mathbf{d}} \leq \operatorname{diam}(\mathsf{n_s}) \leq \alpha_2 \mathbf{N}^{-\mathbf{d}}, \quad \mathbf{s = 1, 2, \ldots, N}. \tag{179}$$

The second means that every cell n_s of $\mathfrak{T}_h$ can be partitioned into n_s regularly shaped triangles ($\mathbf{d = 2}$), or tetrahedrons ($\mathbf{d = 3}$), and $\max\limits_{\mathbf{s}} \mathbf{n_s} \leq \text{const}$.

We analyze samplings where the degrees of freedom represent normal components of the vector field on the edges (interfaces) between cells/subdomains.

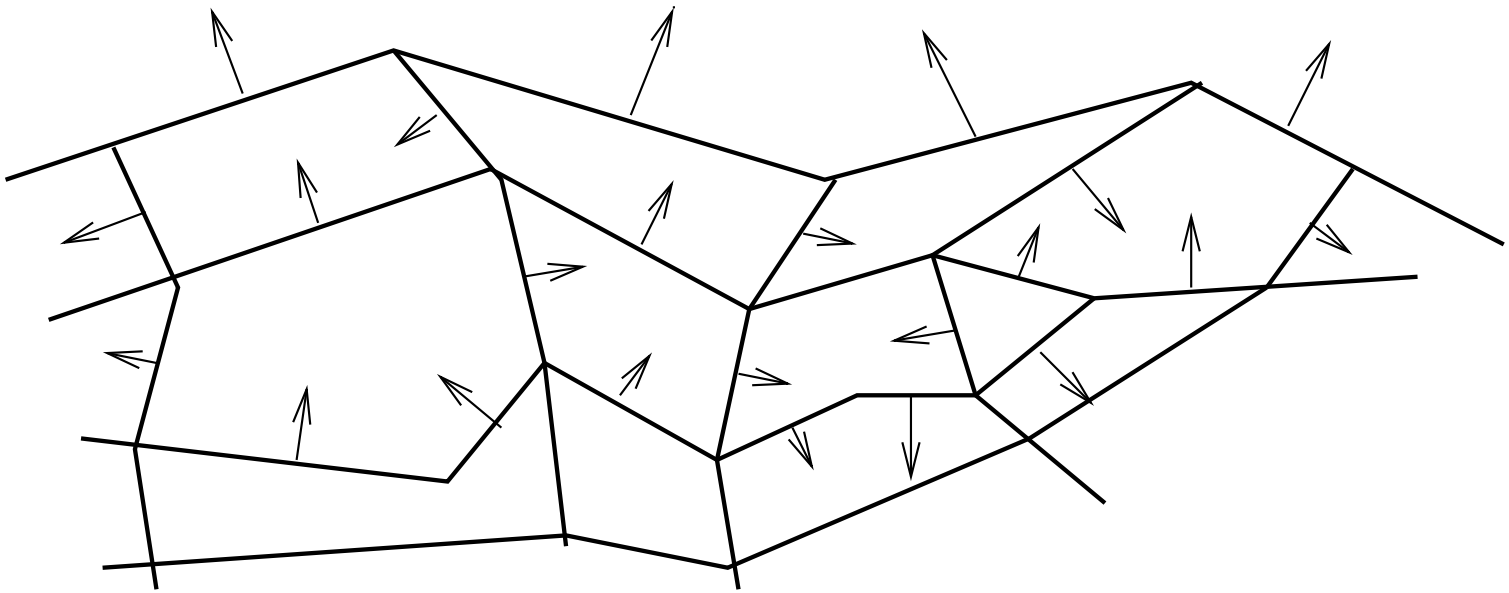
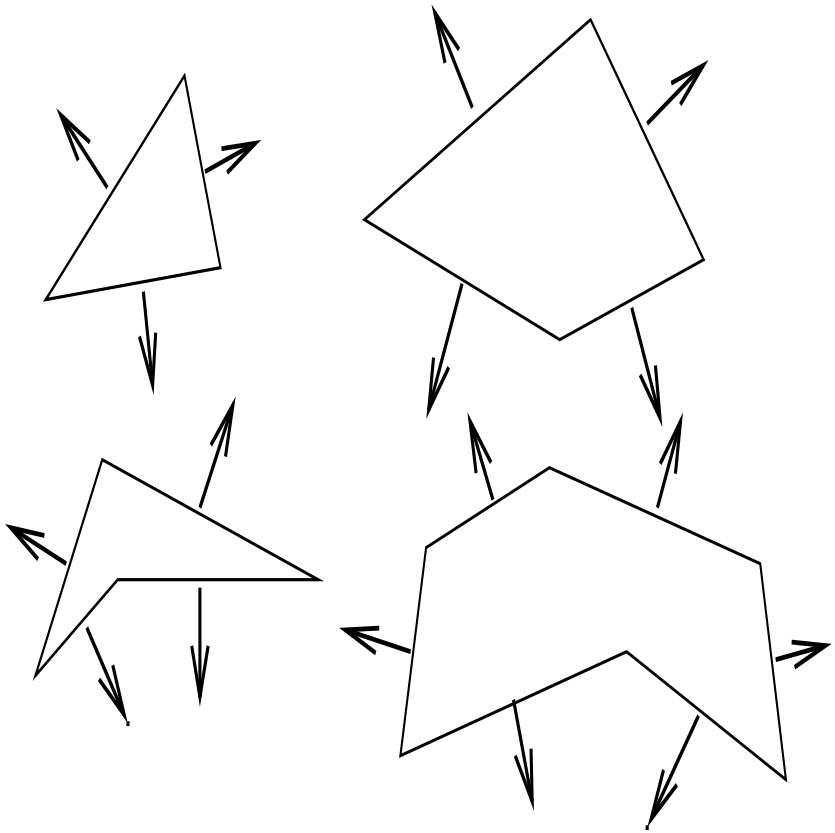
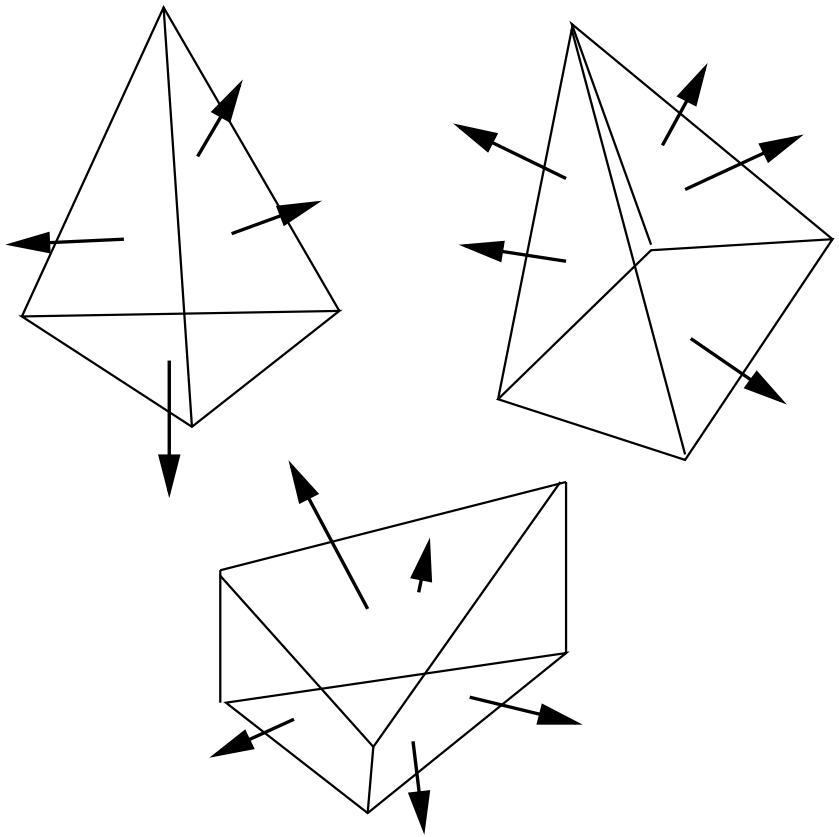


Figure: Distorted polygonal mesh

Elements of a distorted mesh in 2D



Elements of a distorted mesh in 2D



Denote by $\mathbf{E_{st}}$ the common edge of subdomains Π_s and Π_t and by ν_{st} the unit normal vector to this edge oriented such that it is external to Π_s if $s < t$ and external to Π_t in the opposite case. The set of all edges we denote by $\mathcal{E}_h$. The field is approximated by the quantities q_{st} defined on each $\mathbf{E_{st}} \in \mathcal{E}_\mathbf{h}$ (see Figure 1). The value of q_{st} represents the normal component of the vector field on the edge $\mathbf{E_{st}}$. All these quantities form the set

$$\mathbf{q}(\mathfrak{T}_\mathbf{h}) := \{\mathbf{q_{st}} \mid\mid \mathbf{q_{st}} \text{ is defined on } \mathbf{E_{st}} \in \mathcal{E}_\mathbf{h}\}.$$

Besides, for each Π_s we define a number $\mathbf{V_s}$. These numbers form a piecewise constant approximation of the scalar–valued function $\mathbf{v_h}$, i.e.

$$\mathbf{v_h} \in \mathbf{V_h} := \left\{\mathbf{v} \in \mathbf{V}(\Omega) \mid\mid \mathbf{v} \in \mathbf{P^0}(\Pi_\mathbf{s}) \text{ for any } \Pi_\mathbf{s}\right\},$$

where P^0 denotes the set of zero order polynomials.

In general, there are various finite dimensional formulations that can be established on the basis of this set of parameters. We consider the one based on the "conformity concept" and certain extension operators that transform $q \in \mathbf{q}(\mathcal{T}_h)$ to a function in $H(\Omega, \mathrm{div})$. For this purpose, we construct a linear continuous mapping $\mathbb{P}_\Omega : q \to \mathbb{P}_\Omega\, q$ such that

$$\mathcal{Q} = \mathbb{P}_{\mathbf{\Omega}}\mathbf{q} \in \mathbf{H}(\mathbf{\Omega}, \mathrm{div}) \qquad \forall \mathbf{q} \in \mathbf{q}(\mathcal{T}_\mathbf{h}).$$

Then, the respective finite dimensional scheme readily follows from the functional formulation, if the space $Q(\Omega)$ is replaced by the image space of the operator $\mathbb{P}_\Omega$, which we denote by $Q_h(\Omega)$.

To define $\mathbb{P}_\Omega$ we need first to present a suitable way of extension for a cell $\mathsf{\Pi}$. Let

$$\mathbf{q}(\mathsf{\Pi}) := \{\mathbf{q_{st}} \;||\; \mathbf{q_{st}} \text{ is defined on } \mathbf{E_{st}} \in \partial\mathsf{\Pi}\}$$

be the set of normal components of the flux on the edges of $\mathsf{\Pi}$.

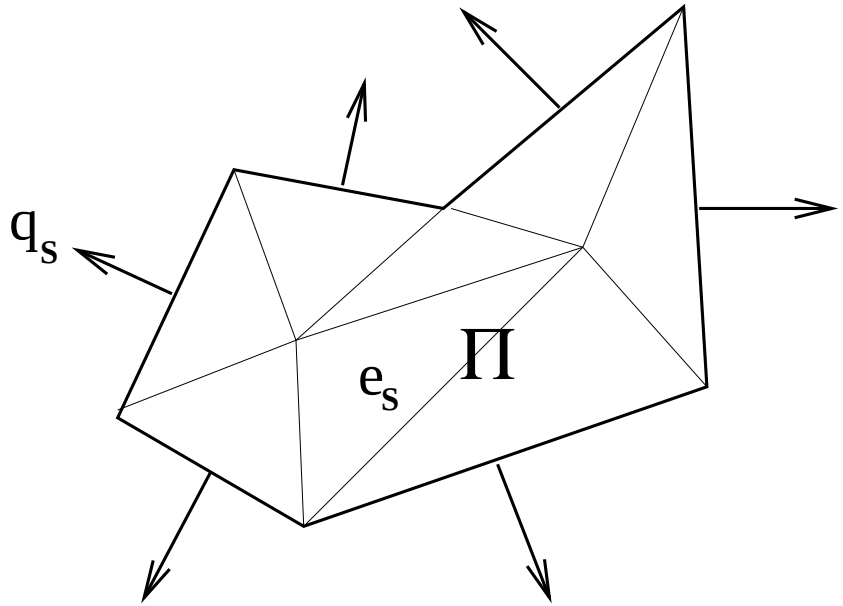


Figure: Polygonal cell with normal components of the flux

On $\mathsf{п}$ we define the *cell extension operator*

$$\mathbb{P}_{\mathsf{п}} : \mathbf{q}(\mathsf{п}) \to \mathbf{H}(\mathsf{п}, \mathrm{div}),$$

which maps the set of normal components of the flux given on $\partial\mathsf{п}$ to a function $\mathcal{Q}_{\mathsf{п}} = \mathbb{P}_{\mathsf{п}}\mathbf{q}(\mathsf{п})$ defined on $\mathsf{п}$. This operator must be linear and satisfy the following conditions:

$$\begin{aligned}
&(\mathbf{a}) \quad \mathrm{div}\,\mathcal{Q}_{\mathsf{п}} \in \mathbf{P}^{\mathbf{0}}(\mathsf{п}), \\
&(\mathbf{b}) \quad \mathcal{Q}_{\mathsf{п}} \text{ is piecevise affine on } \mathsf{п}, \\
&(\mathbf{c}) \quad \int_{\mathsf{п}} \mathrm{div}\,\mathcal{Q}_{\mathsf{п}} \mathbf{dx} = \sum_{\mathbf{E_{st}} \in \partial\mathsf{п}} \mathbf{q_{st}} |\mathbf{E_{st}}|
\end{aligned} \tag{180}$$

Comment: why special Q_h and V_h should be used?

A cell may have a complicated form, but it is "small" so that on each sell we can approximate the pressure $\mathbf{v}$ by a constant, i.e., the first space is

$$\mathbf{V_h} := \{\mathbf{v_h} \in \mathbf{V} = \mathbf{L^2}(\Omega) \mid \mathbf{v_h} \in \mathbf{P^0}(\mathbf{E^i})\},$$

How then $q_h \in H(\Omega, \mathrm{div})$ should be presented on a distorted cell ?
We use the method suggested in Yu. Kuznetsov, S. Repin, JNM, (2003).

It is based on the condition

$$\mathbf{div q_h = const} \qquad \text{on } \mathbf{E^i}$$

Motivation

1. Approximations $\mathbf{q_h}$ and $\mathbf{v_h}$ are compatible in the sense that $\mathbf{div q_h} \in \mathbf{V_h}$. This fact is very important in the stability and convergence analysis.
2. Also, we have approximation arguments to justify such a choice.

Indeed, since $\mathbf{v_h} = \mathbf{const}$ on $\mathbf{E^i}$, the equation

$$\int_\Omega (\mathbf{div p_h} + \mathbf{f})\mathbf{v_h} \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{v_h} \in \mathbf{V_h}.$$

is equivalent to

$$\int_\Omega (\mathbf{div p_h} + [\mathbf{f}]_\mathbf{E})\mathbf{v_h} \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{v_h} \in \mathbf{V_h}.$$

where $[\mathbf{f}]_\mathbf{E}$ is the function taking mean values of $\mathbf{f}$ on each cell.

Thus, in this lowest order approximations, the variability of f inside a cell is ignored and in reality we use the relation

$$\mathbf{div p} + [\mathbf{f}]_{\mathbf{E}} = \mathbf{0}.$$

Therefore, for the flux approximation it is natural to impose the condition:

$$\mathbf{div p_h = const} \qquad \text{on } \mathbf{E^i}$$

Certainly, $\mathbb{P}_\Pi$ must additionally satisfy the natural condition on that the norm $\|\mathcal{Q}_\Pi\|_\Pi$ must be equivalent to the norm of the vector $\mathbf{q}(\Pi)$. This leads to certain requirements on the structure of the mesh, which are satisfied under the assumptions we have imposed on $\mathcal{T}_h$.
Having $\mathbb{P}_{\Pi_s}$ for each cell Π_s, we define the *global extension operator*

$$\mathbb{P}_\Omega : \mathbf{q}(\mathcal{T}_h) \to \mathbf{Q}_h$$

by setting

$$\mathbb{P}_\Omega \mathbf{q}(\mathbf{x}) = \mathbb{P}_{\Pi_s} \mathbf{q}(\Pi_s)(\mathbf{x}), \quad \text{if } \mathbf{x} \in \Pi_s$$

for any $\mathbf{q} \in \mathbf{q}(\mathcal{T}_h)$.

Consider the respective saddle point problem for the Lagrangian

$$\mathbf{L}(\mathbf{v_h}, \mathbf{q_h}) = -\int_{\Omega} \left(\mathbf{v_h} \mathrm{div} \mathbf{q_h} + \frac{\mathbf{1}}{\mathbf{2}} |\mathbf{q_h}|^2 + \mathbf{f} \mathbf{v_h} \right) \mathbf{dx}$$

on $\mathbf{V_h} \times \mathbf{Q_h}$. Components of the saddle point $(\mathbf{u_h}, \mathbf{p_h})$ are solutions of two variational problems associated with the Lagrangian L. The first problem $\mathcal{P}_h$ is to find $u_h \in V_h(\Omega)$ such that

$$\inf_{\mathbf{v_h} \in \mathbf{V_h}(\Omega)} \mathbf{J_h}(\mathbf{v_h}) = \mathbf{J_h}(\mathbf{u_h}) := \inf \mathcal{P}_\mathbf{h},$$

where

$$\mathbf{J_h}(\mathbf{v_h}) = \sup_{\mathbf{q_h} \in \mathbf{Q_h}} \mathbf{L}(\mathbf{v_h}, \mathbf{q_h}).$$

Another (dual) problem $\mathcal{P}_h^*$ is to find p_h that maximizes the functional

$$\mathbf{I}(\mathbf{q_h}) = -\frac{\mathbf{1}}{\mathbf{2}} \int_{\Omega} |\mathbf{q_h}|^2 \, \mathbf{dx}$$

on the set

$$Q_{fh}(\Omega) := \left\{ q_h \in Q_h \; || \int_{\Omega} (\mathrm{div} q_h + f) \mathbf{v_h} \mathbf{dx} = \mathbf{0} \; \forall \mathbf{v_h} \in \mathbf{V_h} \right\}.$$

Well–posedness of the discrete problems

Under the above made assumptions on the external data and $\mathfrak{T}_h$ both these problems $\mathcal{P}_h$ and $\mathcal{P}_h^*$ are stable and well–posed.

Assumption. *For any* $\mathbf{v_h} \in \mathbf{V_h(\Omega)}$ *one can find a function* $\eta_h \in Q_h(\Omega)$ *such that*

$$\mathrm{div}\eta_{\mathbf{h}}|_{\Pi_{\mathbf{s}}} = \mathbf{v_h}|_{\Pi_{\mathbf{s}}}, \quad \mathbf{s} = \mathbf{1}, \mathbf{2}, ...\mathbf{N} \tag{181}$$

$$\|\eta_{\mathbf{h}}\|_{\mathbf{H}(\Omega,\mathrm{div})} \leq \mathbf{C}\|\mathbf{v_h}\|_{\Omega}, \tag{182}$$

where the constant $\mathbf{C}$ *depends on* Ω.

Proposition 1. Functional J_h is coercive on V_h.

Proof

Let $\mathbf{v_h} \in \mathbf{V_h}$. For any $\mathbf{v_h}$, we find η_h in accordance with conditions of Proposition 1. For any positive α

$$\mathbf{J_h(v_h)} = \sup_{\mathbf{q_h} \in \mathbf{Q_h}} \mathbf{L(v_h, q_h)} \geq$$

$$\geq \mathbf{L(v_h, -\alpha \eta_h)} = \int_{\Omega} \left(\alpha \mathbf{v_h} \mathrm{div} \eta_{\mathbf{h}} - \frac{\alpha^2}{2} |\eta_{\mathbf{h}}|^2 - \mathbf{f v_h} \right) \mathbf{dx}.$$

By (181) and (182), we obtain

$$\mathbf{J_h(v_h)} \geq \sum_{\mathbf{s}=1}^{\mathbf{N}} \alpha \|\mathbf{v_h}\|^2_{\mathbf{\Pi_s}} - \frac{\alpha^2 \mathbf{C}^2}{2} \|\mathbf{v_h}\|^2_{\Omega} - \int_{\Omega} \mathbf{f v_h dx} =$$

$$= \alpha \left(1 - \frac{\alpha \mathbf{C}^2}{2} \right) \|\mathbf{v_h}\|^2_{\Omega} - \int_{\Omega} \mathbf{f v_h dx}.$$

Take $\alpha = \frac{1}{C^2}$. Then

$$J_h(v_h) \geq \frac{1}{2C^2}\|v_h\|_\Omega^2 - \int_\Omega f v_h dx \geq$$

$$\geq \frac{1}{4C^2}\|v_h\|_\Omega^2 - C^2\|f\|_\Omega^2 \qquad (183)$$

and the coercivity of J_h on V_h follows.

Proposition 2. J_h is convex and lower semicontinuous on $\mathbf{V_h}$.
Corollary. Since the functional J is convex, lower semicontinuous and coercive on $\mathbf{V_h}$, known theorems of the calculus of variations guarantee that the minimizer u_h exists (see, e.g., [**?**]) .
Moreover, from (183) it follows that

$$\frac{1}{4C^2}\|\mathbf{u_h}\|^2_\Omega \le C^2\|\mathbf{f}\|^2_\Omega + \mathbf{J_h(v_h)} \quad \forall \mathbf{v_h} \in \mathbf{V_h}. \tag{184}$$

For any partition $\mathfrak{T}_h$ the zero function belongs to $\mathbf{V_h}$ and $\mathbf{J_h(0)} = \mathbf{0}$. Therefore, the second term in the right–hand side of (184) vanishes and we see that the norm of $\mathbf{u_h}$ is uniformly bounded with respect to h. This means that the problems $\mathcal{P}_h$ are stable.

It is also not difficult to justify the well–posedness of the dual problem. Indeed, Problem $\mathcal{P}_h^*$ is a problem of maximization a quadratic functional on the affine set Q_{fh}, which has a unique solution provided that this set is not empty.

Since

$$\int_{\Omega} (\mathrm{div}\mathbf{p_h} + \mathbf{f})\mathbf{w_h}\,\mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{w_h} \in \mathbf{V_h}(\Omega), \tag{185}$$

$$\int_{\Omega} (\mathbf{p_h} \cdot \mathbf{q_h} + \mathbf{u_h}\mathrm{div}\mathbf{q_h})\,\mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{q_h} \in \mathbf{Q_h}(\Omega), \tag{186}$$

we arrive at the conclusion that

$$\|\mathbf{p_h}\|_{\Omega}^2 = \int_{\Omega} \mathbf{f}\mathbf{u_h}\,\mathbf{dx} \leq \|\mathbf{f}\|_{\Omega}\|\mathbf{u_h}\|_{\Omega}.$$

In view of (184), we obtain

$$\|\mathbf{p_h}\|^2_{\Omega} \leq \mathbf{2C^2}\|\mathbf{f}\|^2_{\Omega}. \tag{187}$$

It is worth noting that elements of $\mathbf{Q_{fh}(\Omega)}$ satisfy the condition

$$-\left(\mathrm{div}\mathbf{q_h}\right)_{\mathbf{\Pi_s}} = \frac{\mathbf{1}}{|\mathbf{\Pi_s}|}\int\limits_{\mathbf{\Pi_s}} \mathbf{f}\mathbf{dx}.$$

Really, is set $w_h = 0$ on all the cells except Π_i and $w_h = c$ on that single sell. then

$$\int\limits_{\mathbf{\Omega}} (\mathrm{div}\mathbf{q_h} + \mathbf{f})\mathbf{c}\mathbf{dx} = \mathbf{0} \Rightarrow \left[\mathrm{div}\mathbf{q_h} + \mathbf{f}\right]_{\mathbf{\Pi_i}} = \mathbf{0},$$

i.e.,

$$(\mathrm{div}q_h)_{\Pi_s}|\Pi_s| + \int\limits_{\Pi_s} f dx = 0$$

Since $p_h \in Q_{fh}(\Omega)$, we see that

$$(\mathrm{div}\mathbf{p_h})_{\Pi_s} = -\frac{\mathbf{1}}{|\Pi_s|}\int\limits_{\Pi_s} \mathbf{f}\mathbf{dx} = \frac{\mathbf{1}}{|\Pi_s|}\int\limits_{\Pi_s} \mathrm{div}\mathbf{p}. \tag{188}$$

Thus, $\mathrm{div}p_h$ **is the cell–averaging of** $\mathrm{div}p$**.**

From (187) and (188) it follows that

$$\|\mathbf{p_h}\|_{\mathrm{div},\Omega} \leq \mathbf{const}.$$

Moreover, by (188) and the property of averaged functions we conclude that if f is a smooth (or piecewise smooth) function then $\mathrm{div} p_h \rightarrow \mathrm{div} p$ a. e. in Ω.
If $f \in W_2^1(\Omega)$, then from the Poincaré inequality for the functions with zero mean we also find that

$$\|\mathrm{div} p_h - \mathrm{div} p\|_\Omega \rightarrow 0 \text{ as } h \rightarrow 0.$$

Other results on the convergence of $\mathbf{u_h}$ to $\mathbf{u}$ and $\mathbf{p_h}$ to $\mathbf{p}$ follow from the above stability estimates provided that the spaces $\mathbf{V_h}$ and $\mathbf{Q_h}$ are limit dense in $\mathbf{V}$ and $\mathbf{Q}$, respectively: *for any* $\mathbf{V} \in \mathbf{V}$ *(resp.* $\mathbf{q} \in \mathbf{Q}$*) and any positive* ε *one can find* h_ε *such that for* $h \le h_\varepsilon$ $\mathbf{V_h}$ *contains a function* $\mathbf{V}_\varepsilon$ *satisfying the relation* $\|\mathbf{v} - \mathbf{v}_\varepsilon\|_\Omega \le \varepsilon$ *(resp.* $\mathbf{Q_h}$ *contains* q_ε *satisfying the relation* $\|\mathbf{q} - \mathbf{q}_\varepsilon\|_{\mathrm{div},\Omega} \le \varepsilon$*).*

Then, for any $\mathbf{w} \in \mathbf{V}$ and $\mathbf{q} \in \mathbf{Q}$ there exist sequences $\{\mathbf{w_h}\} \in \mathbf{V_h}$ and $\{\mathbf{q_h}\} \in \mathbf{Q_h}$ that **strongly converge** to vw and $\mathbf{q}$ in $\mathbf{V}$ and $\mathbf{Q}$, respectively.

Since the sequences $\{\mathbf{u_h}\}$ and $\{\mathbf{p_h}\}$ are uniformly bounded, they contain subsequences weakly converging to some limits $\widetilde{\mathbf{u}} \in \mathbf{V}$ and $\widetilde{\mathbf{q}} \in \mathbf{Q}$, respectively. Passing to the limit in (185)–(186) we obtain

$$\int_\Omega (\mathrm{div}\widetilde{\mathbf{p}} + \mathbf{f})\mathbf{w}\,\mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{w} \in \mathbf{V}(\Omega), \tag{189}$$

$$\int_\Omega (\widetilde{\mathbf{p}} \cdot \mathbf{q} + \widetilde{\mathbf{u}}\mathrm{div}\mathbf{q})\,\mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{q} \in \mathbf{Q}(\Omega), \tag{190}$$

what means that $\widetilde{\mathbf{u}} = \mathbf{u}$ and $\widetilde{\mathbf{p}} = \mathbf{p}$, i.e. the sequences converge to the solution of the basic problem.

Take sequences $\hat{\mathbf{u}}_\mathbf{h} \rightarrow \mathbf{u}$ in $\mathbf{V}$ and $\hat{\mathbf{p}}_\mathbf{h} \rightarrow \mathbf{p}$ in $\mathbf{Q}$. We have

$$\mathbf{L}(\mathbf{u_h}, \mathbf{q_h}) \leq \mathbf{L}(\mathbf{u_h}, \mathbf{p_h}) \leq \mathbf{L}(\mathbf{v_h}, \mathbf{p_h}) \quad \forall \mathbf{v_h} \in \mathbf{V_h},\ \mathbf{q_h} \in \mathbf{Q_h}. \tag{191}$$

Therefore,

$$\mathbf{J_h}(\mathbf{u_h}) = \sup_{\mathbf{q_h} \in \mathbf{Q_h}} \mathbf{L}(\mathbf{u_h}, \mathbf{q_h}) = \mathbf{L}(\mathbf{u_h}, \mathbf{p_h}), \tag{192}$$

$$\mathbf{I}(\mathbf{p_h}) = \inf_{\mathbf{v_h} \in \mathbf{v_h}} \mathbf{L}(\mathbf{v_h}, \mathbf{p_h}) = \mathbf{L}(\mathbf{u_h}, \mathbf{p_h}) \tag{193}$$

and

$$\lim_{h \rightarrow 0} I(p_h) \geq \lim_{h \rightarrow 0} L(\hat{p}_h, u_h) =$$

$$= \lim_{h \rightarrow 0} \int_\Omega \left(u_h \mathrm{div} \hat{p}_h - \frac{1}{2} |\hat{p}_h|^2 - f u_h \right) dx = L(u, p) = I(p). \tag{194}$$

Since $\mathbf{p_h}$ weakly converges to $\mathbf{p}$, we have

$$\lim_{\mathbf{h}\to\mathbf{0}} \|\mathbf{p_h}\|_\Omega \geq \|\mathbf{p}\|_\Omega,$$

what gives the relation

$$\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{I}(\mathbf{p_h}) \leq \mathbf{I}(\mathbf{p}). \tag{195}$$

From (194) and (195) we conclude that

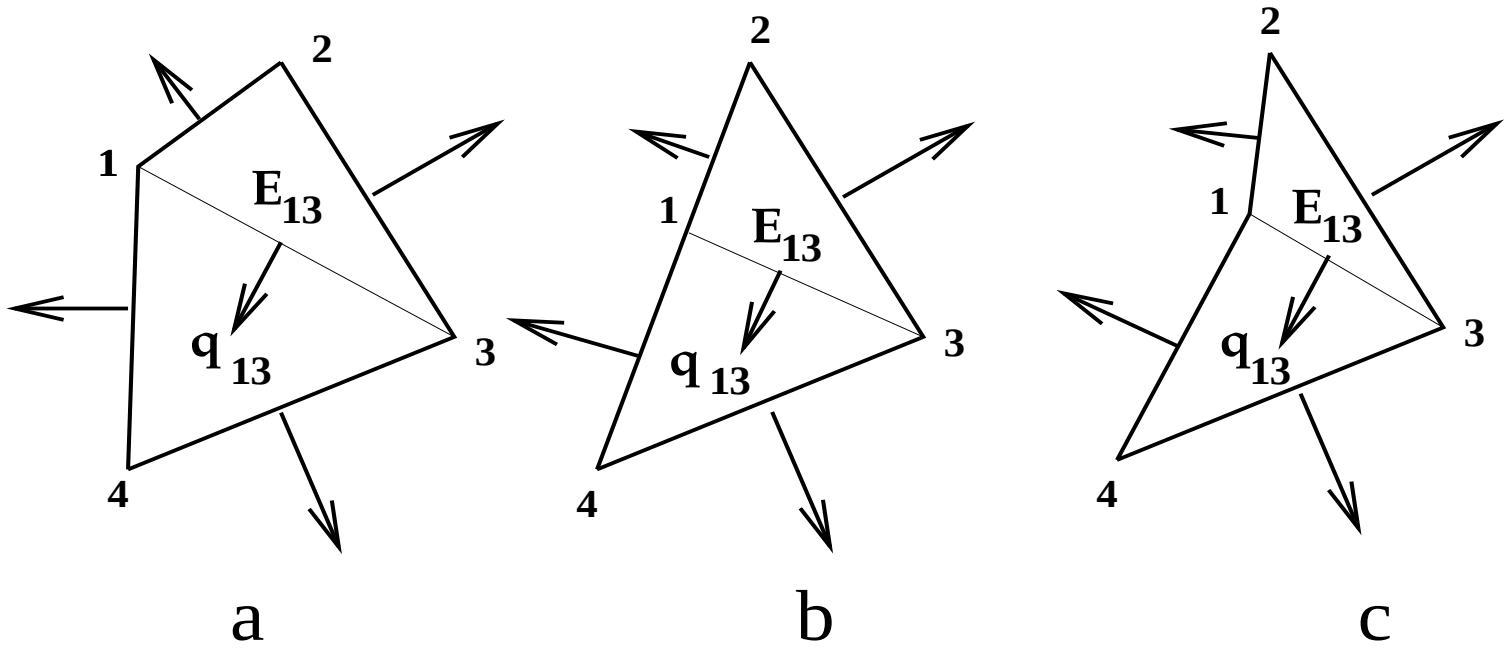
$$\lim_{\mathbf{h}\to\mathbf{0}} \mathbf{I}(\mathbf{p_h}) = \mathbf{I}(\mathbf{p}). \tag{196}$$

Thus, $\|\mathbf{p_h}\|_\Omega \to \|\mathbf{p}\|_\Omega$ and, consequently, $\|\mathbf{p_h} - \mathbf{p}\|_\Omega \to \mathbf{0}$.

Examples of extension operators

Quadrilateral cells in $\mathbb{R}^2$.

Quadrilateral cell $\mathsf{\Pi}$ with nodes numerated 1,2,3, and 4. Edges are denoted by $\mathbf{E_{12}}$, $\mathbf{E_{23}}$, $\mathbf{E_{34}}$, and $\mathbf{E_{14}}$. Below we construct a cell extension operator $\mathbb{P}_{\mathsf{\Pi}}$ valid for a convex, nonconvex and degenerate cells.



To construct an extended function $\mathcal{Q}_{\Pi}$, that satisfies the conditions $\mathcal{Q}_{\Pi}|_{\mathbf{E_{st}}} \cdot \nu_{\mathbf{st}} = \mathbf{q_{st}}$ (where ν_{st} denotes the unit normal to $\mathbf{E_{st}}$) we introduce a subsidiary edge $\mathbf{E_{13}}$ and define the respective normal component $\mathbf{q_{13}}$ from the condition

$$\frac{1}{|\mathbf{T_{123}}|}\int_{\mathbf{T_{123}}} \mathrm{div}(\mathcal{Q}_{\Pi})\mathbf{dx} = \frac{1}{|\mathbf{T_{134}}|}\int_{\mathbf{T_{134}}} \mathrm{div}(\mathcal{Q}_{\Pi})\mathbf{dx}, \quad (197)$$

which means that the divergence of the extended function $\mathcal{Q}_{\Pi}$ is constant on Π.

From (197) we obtain

$$\mathbf{q_{13}} = \rho \left(\frac{\mathbf{q_{12}}|\mathbf{E_{12}}|}{|\mathbf{T_{123}}||\mathbf{E_{13}}|} + \frac{\mathbf{q_{23}}|\mathbf{E_{23}}|}{|\mathbf{T_{123}}||\mathbf{E_{13}}|} + \frac{\mathbf{q_{14}}|\mathbf{E_{14}}|}{|\mathbf{T_{134}}||\mathbf{E_{13}}|} - \frac{\mathbf{q_{34}}|\mathbf{E_{34}}|}{|\mathbf{T_{134}}||\mathbf{E_{13}}|} \right), \quad (198)$$

where

$$\rho = \frac{|\mathbf{T_{123}}||\mathbf{T_{134}}|}{|\mathbf{T_{123}}| + |\mathbf{T_{134}}|}.$$

Now, for each of the two triangles we can construct the field using the lowest order Raviart–Thomas elements. This field meets the conditions $\mathrm{div}\,\mathcal{Q}_{\mathbf{\Pi}} = \mathbf{const}$ in $\mathbf{\Pi}$ and

$$\int_{\mathbf{\Pi}} \mathrm{div}\,\mathcal{Q}_{\mathbf{\Pi}}\, \mathbf{dx} = \mathbf{q_{12}}|\mathbf{E_{12}}| + \mathbf{q_{23}}|\mathbf{E_{23}}| + \mathbf{q_{34}}|\mathbf{E_{34}}| + \mathbf{q_{14}}|\mathbf{E_{14}}|. \quad (199)$$

It is possible to show that the extension operator satisfies the relation

$$\|\mathcal{Q}_{\Pi}\|^2_{\Pi} \leq \mathbf{h}^2 \mu_2(\Pi) \sum_{\mathbf{E_{st}} \in \partial \Pi} \mathbf{q}^2_{\mathbf{st}}, \tag{200}$$

where μ_2 is a positive constant depending on the shape of a cell Π. For this purpose, at each node, we define vectors associated with certain triangle containing this node. It means that it is bounded.

Polygonal cells in $\mathbb{R}^3$

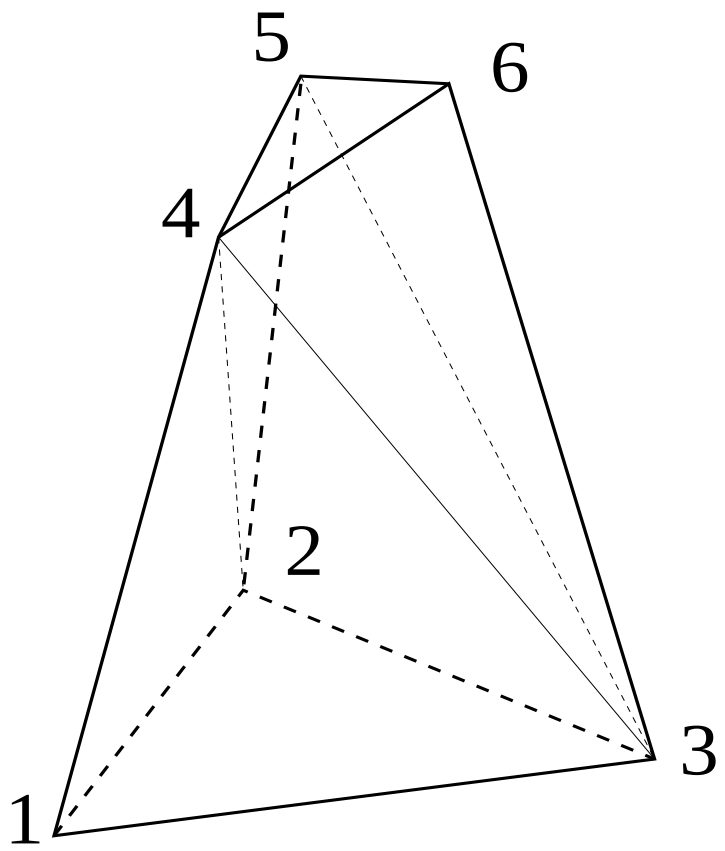


Figure: Extension inside a prism

Denote the normal components of the flux on the faces $\mathbf{E_{1245}}$, $\mathbf{E_{2356}}$, $\mathbf{E_{1346}}$, $\mathbf{E_{123}}$, and $\mathbf{E_{456}}$ by q_1, q_2, q_3, q_4, and q_5, respectively. Introduce two subsidiary internal faces $\mathbf{E_{234}}$ and $\mathbf{E_{345}}$ with normal components q' and q'' oriented inside $\mathbf{\Pi}$. Now, the prism is divided into 3 tetrahedrons T_{1234}, T_{2345}, and T_{3456}. The divergence of the extended field is defined via the Stokes theorem as

$$\mathrm{div}\mathbb{P}_{\mathbf{\Pi}} = \mathbf{g}_{\mathbf{\Pi}},$$

where

$$\mathbf{g}_{\mathbf{\Pi}} = \mathbf{q_1}|\mathbf{E_{1245}}| + \mathbf{q_2}|\mathbf{E_{2356}}| + \mathbf{q_3}|\mathbf{E_{1346}}| + \mathbf{q_4}|\mathbf{E_{123}}| + \mathbf{q_5}|\mathbf{E_{456}}|.$$

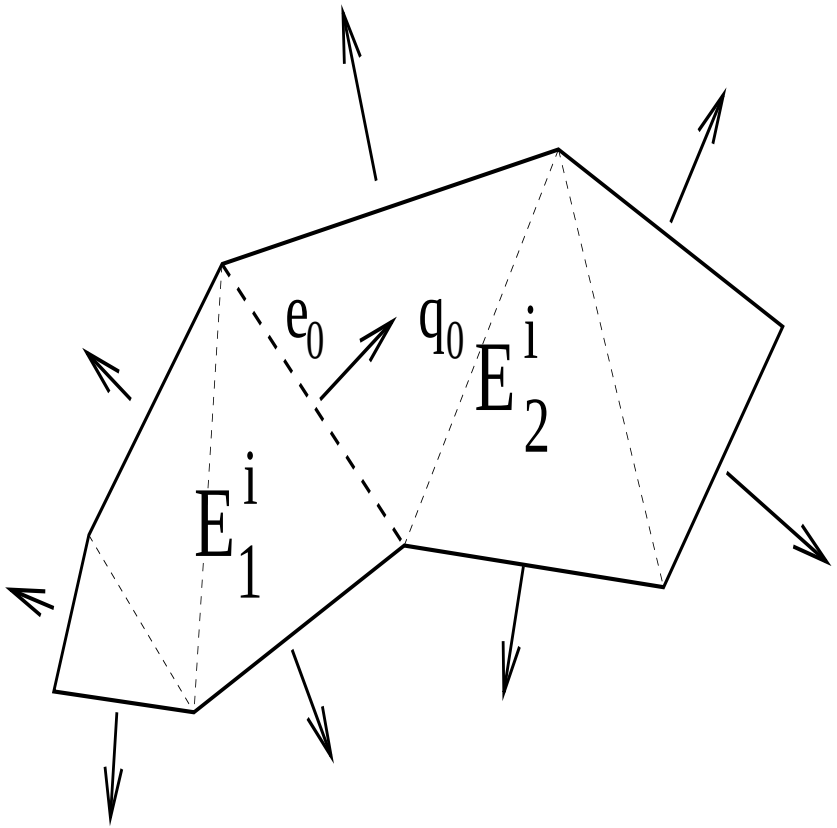
Since the condition $\mathrm{div}\mathbb{P}_{\Pi} = \mathbf{g}_{\Pi}$ must hold on each of the tetrahedrons, we obtain

$$\mathbf{q}_1|\mathbf{E}_{124}| + \mathbf{q}'|\mathbf{E}_{234}| + \mathbf{q}_3|\mathbf{E}_{134}| + \mathbf{q}_4|\mathbf{E}_{123}| = \mathbf{g}_{\Pi}|\mathbf{T}_{1234}|,$$
$$\mathbf{q}_2|\mathbf{E}_{356}| + \mathbf{q}''|\mathbf{E}_{345}| + \mathbf{q}_3|\mathbf{E}_{346}| + \mathbf{q}_5|\mathbf{E}_{456}| = \mathbf{g}_{\Pi}|\mathbf{T}_{3456}|.$$

Thus, the numbers q' and q'' and the respective extension $\mathcal{Q}_{\Pi}$ are uniquely defined.

General method of constructing $\mathbb{P}_\Pi$

In more complicated cases as, e.g., for



internal fluxes are easily excluded by the $\operatorname{div} q_h = \text{const}$ condition and the Stokes theorem.

In fact, we have constructed an interpolation operator π^Q on each cell.
If $\mathbf{q}$ is a vector–valued function defined on $\mathbf{E^i}$ having summable traces, then the normal flux on $\mathbf{e_{ij}} \subset \partial \mathbf{E^i}$ can be defined as follows

$$\mathbf{q^{ij}} = \frac{\mathbf{1}}{|\mathbf{e_{ij}}} \int\limits_{\mathbf{e_{ij}}} \mathbf{q} \cdot \nu_{\mathbf{ij}} \mathbf{ds}$$

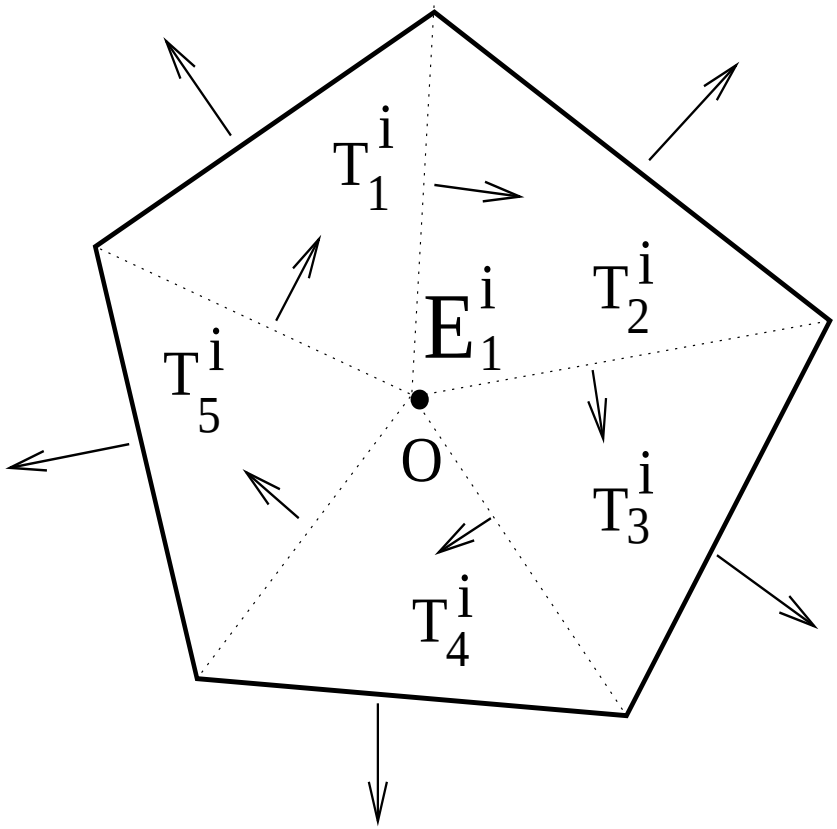
By extending the field inside $\mathbf{E^i}$ we obtain the function $\mathbf{q_h}$ such that

$$\mathbf{q_h} \in \mathbf{H}(\Omega, \mathbf{div}), \quad \mathbf{div q_h} = \mathbf{const} \text{ on } \mathbf{E^i}$$
$$\mathbf{q_h} \text{ is piecewise affine in } \mathbf{E^i}$$

and, thus, define the interpolation operator

$$\pi^{\mathbf{Q}} : \mathcal{Q} \to \mathbf{Q_h}.$$

A cell with internal nodes



For such a sell there are many different fields that satisfy the condition

$$\mathbf{divq_h} = \text{const.}$$

To avoid nonuniqueness the condition

$$\|\mathbf{q_h}\|_{\mathbf{E^i}} \rightarrow \min$$

should be used.

We may *equivalently formulate additional condition* as follows.
Using the lowest order Raviart–Thomas finite elements we find "local" $\mathcal{Q}_\Pi$ satisfying the boundary condition $\mathcal{Q}_\Pi \cdot \nu = \mathbf{q}(\Pi) \cdot \nu$ as a solution of the discrete problem

$$\int_\Pi (\mathbf{v}_\tau \mathrm{div}\, \mathcal{Q}_\Pi + \mathbf{g}_\Pi \mathbf{v}_\tau) \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{v}_\tau \in \mathbf{V}_\tau, \tag{201}$$

$$\int_\Pi (\mathcal{Q}_\Pi \cdot \mathbf{q}_\tau + \mathbf{u}_\tau \mathrm{div}\, \mathbf{q}_\tau) \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{q}_\tau \in \mathbf{Q_{0\tau}}, \tag{202}$$

where $\mathbf{V}_\tau$ is the space of piecewise constant functions and $Q_{0\tau}$ is the finite element space formed by Raviart–Thomas elements contained in $H(\Omega, \mathrm{div})$ and subject to the boundary condition

$$q_\tau \cdot \nu = 0 \qquad \text{on } \partial\Pi.$$

It is well known that this problem has a unique solution $\mathcal{Q}_\Pi$ and the function u_τ is uniquely defined up to an arbitrary constant.
In virtue of (201),

$$\operatorname{div}\mathcal{Q}_\Pi + g_\Pi = 0 \quad \text{in}\, \Pi\,, \tag{203}$$

so that the requirement $\operatorname{div}\mathcal{Q}_\Pi = const$ is satisfied. Since $\mathcal{Q}_\Pi \in H(\Pi, \operatorname{div})$, the condition (180c) is also satisfied.

Note that the function $\mathcal{Q}_\Pi$ can be also viewed as a solution of the respective discrete problem with Neumann boundary conditions, which is to maximize the functional

$$I_\Pi(q_\tau) = -\frac{1}{2}\int_\Pi |q_\tau|^2 dx$$

on the set of fields satisfying the condition

$$\mathrm{div} q_\tau = g_\Pi$$

and the prescribed boundary conditions . If Π is divided into a few subdomains (as, e.g., in the above example for quadrilaterals), then such a condition may uniquely define the field. However, in other cases fields with constant divergence may create a subspace. In this case, uniqueness of the extension is provided by the fact that the minimizer of the quadratic functional I_Π on this subspace is uniquely defined.

Rate convergence estimates. General line

We analyze 3 different mixed approximations on distorted meshes:
A. The finest approximation on the mesh $\widetilde{\mathcal{T}}_{\mathbf{h}}$ when all cells are decomposed into simplexes. On simplexes the pressure field is assumed to be constant and the flux is approximated by the $\mathbf{RT^0}$ elements. This gives the pair of spaces $(\widetilde{\mathbf{V}}_{\mathbf{h}}, \widetilde{\mathbf{Q}}_{\mathbf{h}})$.
B. The approximation on the mesh $\mathcal{T}_{\mathbf{h}}$ that consists of all cells. On cells the pressure field is assumed to be constant and the flux is approximated by the procedure discussed further. This gives the pair of spaces $(\mathbf{V}_{\mathbf{h}}, \mathbf{Q}_{\mathbf{h}})$.
C. $\mathcal{T}_{\mathbf{h}}$ is the same, but normal fluxes are averaged on each "macroface" $\mathbf{e}$. This gives the pair of spaces $(\mathbf{V}^{\mathbf{e}}_{\mathbf{h}}, \mathbf{Q}^{\mathbf{e}}_{\mathbf{h}})$.

"Referenced" problem

Consider the dual mixed formulations on $\widetilde{\mathcal{T}}_h$:
Problem $\widetilde{\mathcal{P}}_h$: find $(\widetilde{p}_h, \widetilde{u}_h) \in \widetilde{Q}_h \times \widetilde{V}_h$ such that

$$\int_\Omega (\widetilde{p}_h \cdot \widetilde{q}_h + \widetilde{u}_h \mathrm{div} \widetilde{q}_h) dx = 0 \qquad \forall \widetilde{q}_h \in \widetilde{Q}_h$$

$$\int_\Omega (\mathrm{div} \widetilde{p}_h + f) \widetilde{v}_h dx = 0 \qquad \forall \widetilde{v}_h \in \widetilde{V}_h .$$

$$\widetilde{V}_h := \{\widetilde{v}_h \in V = L^2(\Omega) \mid \widetilde{v}_h \in P^0(T_j^i)\},$$

$\widetilde{Q}_h \subset Q = H(\Omega, \mathrm{div})$ is constructed by RT^0 elements on the mesh $\widetilde{\mathcal{T}}_h$.

Properties of Problem $\widetilde{\mathcal{P}}$

Problem $\widetilde{\mathcal{P}}_{\mathbf{h}}$ is used as the "referenced" one. Its properties are known (see, e.g., *F. Brezzi, M. Fortin*, *J.E. Roberts and J.-M. Thomas*, 1991.)

1. If $\widetilde{\mathcal{T}}_{\mathbf{h}}$ is regular then the inf-sup condition

$$\inf_{\widetilde{\mathbf{v}}_{\mathbf{h}} \in \widetilde{\mathbf{V}}_{\mathbf{h}}} \sup_{\widetilde{\mathbf{q}}_{\mathbf{h}} \in \widetilde{\mathbf{Q}}_{\mathbf{h}}} \frac{\int_{\Omega} \widetilde{\mathbf{v}}_{\mathbf{h}} \mathbf{div} \widetilde{\mathbf{q}}_{\mathbf{h}} \, \mathbf{dx}}{\|\widetilde{\mathbf{v}}_{\mathbf{h}}\| \|\widetilde{\mathbf{q}}_{\mathbf{h}}\|_{\mathbf{div},\Omega}} \geq \widetilde{\mathbf{C}} \tag{204}$$

holds with a constant $\mathbf{C}$ independent of $\mathbf{h}$.

2. Problem $\widetilde{\mathcal{P}}_{\mathbf{h}}$ is uniquely solvable for any $\mathbf{h} > \mathbf{0}$ and

$$\widetilde{\mathbf{u}}_{\mathbf{h}} \to \mathbf{u} \text{ in } \mathbf{V}(\Omega), \quad \widetilde{\mathbf{p}}_{\mathbf{h}} \to \mathbf{p} \text{ in } \mathbf{Q}.$$

Moreover, for the $\mathbf{RT}^{\mathbf{0}}$ approximations we have the standard rate convergence estimate

$$\|\widetilde{\mathbf{p}}_{\mathbf{h}} - \mathbf{p}\|_{\mathbf{div},\Omega} + \|\widetilde{\mathbf{u}}_{\mathbf{h}} - \mathbf{u}\| \leq$$
$$\leq \mathbf{C}_{\mathbf{RT}}\mathbf{h}(|\mathbf{u}|_{1,\Omega} + \|\mathbf{p}\|_{1,\Omega} + \|\mathbf{div}\mathbf{p}\|_{1,\Omega})$$

that holds provided that the exact solution is sufficiently regular.

We use the above properties of Problem $\widetilde{\mathcal{P}}_{\mathbf{h}}$ in order to establish similar properties for Problems $\mathcal{P}_{\mathbf{h}}$ and $\mathcal{P}^{\mathbf{e}}_{\mathbf{h}}$.

First dual mixed formulations on $\mathfrak{T}_h$

Problem $\mathcal{P}_h$: find $(\mathbf{p_h}, \mathbf{u_h}) \in \mathbf{Q_h} \times \mathbf{V_h}$ such that

$$\int_\Omega (\mathbf{p_h} \cdot \mathbf{q_h} + \mathbf{u_h} \mathbf{div} \mathbf{q_h}) \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{q_h} \in \mathbf{Q_h} \tag{205}$$

$$\int_\Omega (\mathbf{div} \mathbf{p_h} + \mathbf{f}) \mathbf{v_h} \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{v_h} \in \mathbf{V_h}. \tag{206}$$

$$\mathbf{V_h} := \{\mathbf{v_h} \in \mathbf{V} = \mathbf{L^2}(\Omega) \mid \mathbf{v_h} \in \mathbf{P^0}(\mathbf{E^i}),\ \mathbf{i} = \mathbf{1}, \mathbf{2}, ...\mathbf{N}\},$$

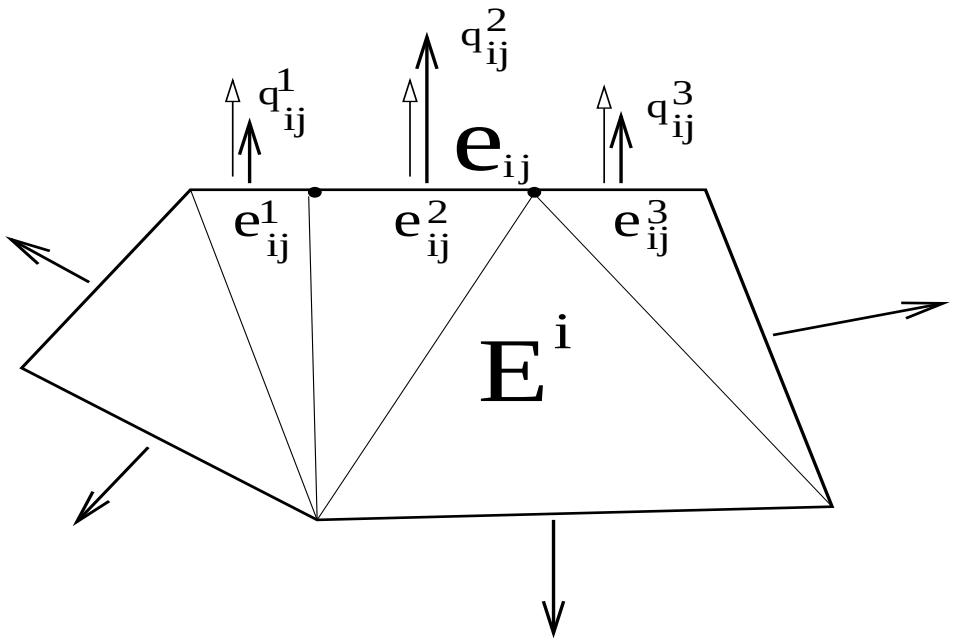
$$\mathbf{Q_h} := \{\mathbf{q_h} \in \mathbf{Q_h} \mid \mathbf{div} \mathbf{q_h} \in \mathbf{P^0}(\mathbf{E^i}),\ \mathbf{i} = \mathbf{1}, \mathbf{2}, ...\mathbf{N}\}$$

Mathematical properties of this problem are in focus of the investigation

Reduced space Q_h^e

In real life problems, analysts are often faced with meshes having "nonmatching" faces.

Certain faces of a cell may belong to one common plane and create a common boundary of one cell (macroface) denoted by Γ_h. We can reduce degrees of freedom on a macroface if replace different normal fluxes given on **e** by one on the basis of the condition:
the value of $\int_e q_h \cdot \nu$ remains unchanged.



A macroface

By this "averaging" procedure we replace $\mathbf{q_h}$ by $\mathbf{q_h^e} \in \mathbf{Q_h^e}$ and obtain a new space for fluxes:

$$\mathbf{Q_h^e} := \{\mathbf{q_h} \in \mathbf{Q_h} \mid \mathbf{q_h} \cdot \nu = \mathbf{const} \quad \forall \mathbf{e} \in \mathbf{\Gamma_h}\}$$

Second dual mixed formulations on $\mathfrak{T}_h$

Problem $\mathcal{P}_h^{\mathbf{e}}$ Find $(\mathbf{p_h^e}, \mathbf{u_h^e}) \in \mathbf{Q_h^e} \times \mathbf{V_h}$ such that

$$\int_\Omega (\mathbf{p_h^e} \cdot \mathbf{q_h^e} + \mathbf{u_h^e div q_h^e}) \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{q_h^e} \in \mathbf{Q_h^e}, \tag{207}$$

$$\int_\Omega (\mathbf{div p_h^e} + \mathbf{f}) \mathbf{v_h dx} = \mathbf{0} \qquad \forall \mathbf{v_h} \in \mathbf{V_h}. \tag{208}$$

It is clear that

$$\mathbf{V_h} \subset \widetilde{\mathbf{V}}_\mathbf{h} \subset \mathbf{V} \qquad \text{and} \qquad \mathbf{Q_h^e} \subset \mathbf{Q_h} \subset \widetilde{\mathbf{Q}} \subset \mathbf{Q}.$$

Main goals

We show that if the above "aggregation–reduction" procedures are properly done and Problem $\widetilde{\mathcal{P}}_h$ is stable and convergent, then

(a) the inf–sup conditions for $(\mathbf{V_h}, \mathbf{Q_h})$ and $(\mathbf{V_h}, \mathbf{Q_h^e})$ are satisfied;

(b) $\mathbf{q_h}$ and $\mathbf{q_h^e}$ converges to p in Q with the same rate as $\widetilde{\mathbf{q}}_\mathbf{h}$ and $\mathbf{v_h}$ and $\mathbf{v_h^e}$ converges to u in V with the same rate as $\widetilde{\mathbf{v}}_\mathbf{h}$;

(c) computational errors for the approximations of all types can be explicitly controlled by the a posteriori estimates presented.

Assumptions

More precisely, we assume that

1. $\mathbf{c_1 h} \le \mathrm{diam}\mathbf{E^i} \le \mathbf{c_2 h}, \qquad \forall \mathbf{E^i} \in \mathfrak{T}_\mathbf{h},$
2. Number of simplexes in a cell is from $\mathbf{1}$ to $\mathbf{m_{max}}$.
3. $\{\widetilde{\mathcal{T}}_h\}$ is regular in the usual sense:
   All angles of simplexes are uniformly bounded,
   Any face of a simplex is either a part of a boundary or a face of another simplex

$$\theta_1 \mathbf{h^d} \le |\mathbf{T^i_j}| \le \theta_2 \mathbf{h^d}, \quad \forall \mathbf{T^i_j} \in \widetilde{\mathcal{T}}_\mathbf{h}$$

If $\mathbf{e}$ is a face of $\mathbf{T^i_j}$ then

$$\gamma_1 \mathbf{h^{d-1}} \le |\mathbf{e}| \le \gamma_2 \mathbf{h^{d-1}}.$$

Logical scheme

1. $\mathbf{Inf\ Sup\ in}\ \widetilde{\mathcal{P}}_{\mathbf{h}} \Rightarrow \mathbf{Inf\ Sup\ in}\ \mathcal{P}_{\mathbf{h}} \Rightarrow \mathbf{Inf\ Sup\ in}\ \mathcal{P}_{\mathbf{h}}^{\mathbf{e}}$
2. **Construct interpolation operators**

$$\pi^{\mathbf{Q}} : \mathcal{Q} \to \mathbf{Q_h} \text{ and } \pi^{\mathbf{V}} : \mathbf{V} \to \mathbf{V_h}$$

3. **Establish projection type estimates**
4. **Obtain estimates for** $\|\widetilde{\pi}^{\mathbf{Q}}\mathbf{p} - \pi^{\mathbf{Q}}\mathbf{p}\|$ and $\|\widetilde{\pi}^{V} p - \pi^{V} p\|$

Inf–sup condition for cell–approximations

First, we prove that for the cell-approximations Inf-Sup condition holds:

Proposition

Let $\{\widetilde{\mathcal{T}}_h\}$ satisfies the above made assumptions. Then, for any $\mathbf{V_h} \in \mathbf{V_h}$ there exists a vector–valued function $\eta_h^v \in Q_h$ such that

$$-\mathbf{div}\eta_{\mathbf{h}}^{\mathbf{v}} = \mathbf{v_h} \quad \text{on each } \mathbf{E^i} \tag{209}$$

$$\|\eta_h^v\|_{\mathbf{div},\Omega} \leq C\|v_h\|, \tag{210}$$

where $\mathbf{C}$ does not depend on h and $\mathbf{V_h}$.

Proof.

Set $\mathbf{f} = \mathbf{v_h}$. Let $(\widetilde{\mathbf{p}}^{\mathbf{v}}, \widetilde{\mathbf{u}}^{\mathbf{v}}) \in \widetilde{\mathbf{Q}}_{\mathbf{h}} \times \widetilde{\mathbf{V}}_{\mathbf{h}}$ be the pair of functions that satisfies the system

$$\int_\Omega (\widetilde{\mathbf{p}}^{\mathbf{v}}_{\mathbf{h}} \cdot \widetilde{\mathbf{q}}_{\mathbf{h}} + \widetilde{\mathbf{u}}^{\mathbf{v}}_{\mathbf{h}} \mathbf{div} \widetilde{\mathbf{q}}_{\mathbf{h}}) \mathbf{dx} = \mathbf{0} \qquad \forall \widetilde{\mathbf{q}}_{\mathbf{h}} \in \widetilde{\mathbf{Q}}_{\mathbf{h}} \tag{211}$$

$$\int_\Omega (\mathbf{div} \widetilde{\mathbf{p}}^{\mathbf{v}}_{\mathbf{h}} + \mathbf{v_h}) \widetilde{\mathbf{v}}_{\mathbf{h}} \mathbf{dx} = \mathbf{0} \qquad \forall \widetilde{\mathbf{v}}_{\mathbf{h}} \in \widetilde{\mathbf{V}}_{\mathbf{h}}. \tag{212}$$

We observe that

$$-\mathbf{div} \widetilde{\mathbf{p}}^{\mathbf{v}}_{\mathbf{h}} = \mathbf{v_h} \quad \text{on any } \mathbf{T}^{\mathbf{i}}_{\mathbf{j}}.$$

Since $\mathbf{v_h}$ is constant on $\mathbf{E^i}$, we arrive at the conclusion that

$$-\mathbf{div}\widetilde{\mathbf{p}}^{\mathbf{v}}_{\mathbf{h}} = \mathbf{v_h} \quad \text{on any } \mathbf{E^i}.$$

Now we construct the required η_h^v by means of $\widetilde{\mathbf{p}}^{\mathbf{v}}_{\mathbf{h}}$ as follows. On each cell we set normal components of η_i^v equal to the normal components of $\widetilde{\mathbf{p}}^{\mathbf{v}}_{\mathbf{i}}$. If $\mathbf{E^i}$ contains no internal points, then (265) uniquely defines η_h^v by the values of normal fluxes on $\partial\mathbf{E^i}$. Thus, in this case, we simply set $\eta_h^v = \widetilde{\mathbf{p}}^{\mathbf{v}}_{\mathbf{h}}$ in the cells.

If $\mathbf{E^i}$ has an internal point, then η_h^v is defined in accordance with *minimal energy principle*, i.e.,

$$\|\eta_h^v\|^2_{\mathbf{E^i}} = \inf_{\substack{\tau^h \in \aleph(\mathbf{E^i}) \\ \mathbf{div}\tau_\mathbf{h} + \mathbf{v_h} = \mathbf{0}}} \|\tau_h\|^2_{\mathbf{E^i}}. \tag{213}$$

Since $\widetilde{\mathbf{p}}_\mathbf{h}^\mathbf{v} \in \aleph(\mathbf{E^i})$ and $\mathbf{div}\widetilde{\mathbf{p}}_\mathbf{h}^\mathbf{v} + \mathbf{v_h} = \mathbf{0}$ we observe that

$$\|\eta_h^v\| \leq \|\widetilde{\mathbf{p}}_\mathbf{h}^\mathbf{v}\|. \tag{214}$$

Set $\widetilde{\mathbf{q}}_h = \widetilde{\mathbf{p}}_h^v$ in (268). Then,

$$\|\widetilde{\mathbf{p}}_h^v\|^2 \le \|\widetilde{\mathbf{u}}_h^v\| \|\mathbf{div}\widetilde{\mathbf{p}}_h^v\| = \|\widetilde{\mathbf{u}}_h^v\| \|\widetilde{\mathbf{v}}_h\|. \tag{215}$$

By assumption Inf–Sup condition for $\widetilde{\mathbf{Q}}_h \times \widetilde{\mathbf{V}}_h$ holds! Therefore, for the function $\widetilde{\mathbf{u}}_h^v$ we can find a vector–valued function $\widetilde{\mathbf{q}}_h^u \in \widetilde{\mathbf{Q}}_h$ such that

$$\int_\Omega \widetilde{\mathbf{u}}_h^v \mathbf{div}\widetilde{\mathbf{q}}_h^u \, \mathbf{dx} \ge \widetilde{\mathbf{C}} \|\widetilde{\mathbf{u}}_h^v\| \|\widetilde{\mathbf{q}}_h^u\|$$

Since

$$\int_\Omega (\widetilde{\mathbf{p}}_h^v \cdot \widetilde{\mathbf{q}}_h^u + \widetilde{\mathbf{u}}_h^v \mathbf{div}\widetilde{\mathbf{q}}_h^u) \mathbf{dx} = \mathbf{0}$$

we observe that

$$\widetilde{\mathbf{C}} \|\widetilde{\mathbf{u}}_h^v\| \|\widetilde{\mathbf{q}}_h^u\| \le -\int_\Omega \widetilde{\mathbf{p}}_h^v \cdot \widetilde{\mathbf{q}}_h^u \, \mathbf{dx} \le \|\widetilde{\mathbf{p}}_h^v\| \|\widetilde{\mathbf{q}}_h^u\|$$

Therefore,

$$\|\widetilde{\mathbf{u}}_{\mathbf{h}}^{\mathbf{v}}\| \leq \frac{\mathbf{1}}{\widetilde{\mathbf{C}}} \|\widetilde{\mathbf{p}}_{\mathbf{h}}^{\mathbf{v}}\|.$$

Now, (215) leads to the estimate

$$\|\widetilde{\mathbf{p}}_{\mathbf{h}}^{\mathbf{v}}\| \leq \frac{\mathbf{1}}{\widetilde{\mathbf{C}}} \|\mathbf{v}_{\mathbf{h}}\|. \tag{216}$$

Since $\eta_h^v = \widetilde{\mathbf{p}}_{\mathbf{h}}^{\mathbf{v}}$, we have

$$\|\eta_{\mathbf{h}}^{\mathbf{v}}\|_{\mathbf{div},\Omega}^{\mathbf{2}} \leq \left(\mathbf{1} + \frac{\mathbf{1}}{\widetilde{\mathbf{C}}^{\mathbf{2}}}\right) \|\mathbf{v}_{\mathbf{h}}\|^{\mathbf{2}}.$$

Corollary 1. It is easy to see that under the conditions of Proposition 1, the discrete inf-sup condition for the spaces $(\mathbf{Q_h}, \mathbf{V_h})$ holds. Indeed, take arbitrary $\mathbf{v_h} \in \mathbf{V_h}$. We have

$$\sup_{\mathbf{q_h} \in \mathbf{Q_h}} \frac{\int_\Omega \mathbf{v_h} \mathbf{div} \mathbf{q_h} \, \mathbf{dx}}{\|\mathbf{v_h}\| \|\mathbf{q_h}\|_{\mathbf{div},\Omega}} \geq \frac{\int_\Omega \mathbf{v_h} \mathbf{div} \eta_{\mathbf{h}}^{\mathbf{v}} \, \mathbf{dx}}{\|\mathbf{v_h}\| \|\eta_{\mathbf{h}}^{\mathbf{v}}\|_{\mathbf{div},\Omega}} \frac{\|\mathbf{v_h}\|}{\|\eta_h^v\|_{\mathbf{div},\Omega}} = C.$$

Thus, for $(\mathbf{V_h}, \mathbf{Q_h})$ the inf-sup condition holds with a certain positive constant $\mathbf{C}$ depending on $\widetilde{\mathbf{C}}$.

If for cell–approximations Inf–Sup holds, then for them PROJECTION ESTIMATE HOLDS

$$\|\mathbf{p}-\mathbf{p_h}\|_{\mathbf{div},\Omega}+\|\mathbf{u}-\mathbf{u_h}\|\leq$$
$$\leq \mathbf{C}\Big\{\inf_{\mathbf{q_h}\in\mathbf{Q_h}}\|\mathbf{p}-\mathbf{q_h}\|_{\mathbf{div},\Omega}++\inf_{\mathbf{w_h}\in\mathbf{V_h}}\|\mathbf{u}-\mathbf{w_h}\|\Big\}.$$

Key point in deriving a priori estimates

In projection estimates we have quantities of the type

$$\inf_{\mathbf{q_h}\in\mathbf{Q_h}} \|\mathbf{p}-\mathbf{q_h}\|, \quad \inf_{\mathbf{v_h}\in\mathbf{V_h}} \|\mathbf{u}-\mathbf{v_h}\|.$$

We estimate them by

$$\begin{aligned}\|\mathbf{p}-\pi^{\mathbf{Q}}\mathbf{p}\| &\le \|\mathbf{p}-\widetilde{\pi}^{\mathbf{Q}}\mathbf{p}\| + \|\widetilde{\pi}^{\mathbf{Q}}\mathbf{p}-\pi^{\mathbf{Q}}\mathbf{p}\|,\\ \|\mathbf{u}-\pi^{\mathbf{V}}\mathbf{u}\| &\le \|\mathbf{u}-\widetilde{\pi}^{\mathbf{V}}\mathbf{u}\| + \|\widetilde{\pi}^{\mathbf{V}}\mathbf{u}-\pi^{\mathbf{V}}\mathbf{u}\|\end{aligned}$$

For $\mathbf{RT^0}$ approximations on $\widetilde{\mathcal{T}}_h$ we have the estimate

$$\|\mathbf{p}-\widetilde{\pi}^{\mathbf{Q}}\mathbf{p}\| \sim \mathbf{h}, \quad \|\mathbf{u}-\widetilde{\pi}^{\mathbf{Q}}\mathbf{u}\| \sim \mathbf{h}.$$

Therefore, we need to estimate the difference between interpolation operators on the "fine" and "coarse" meshes.

Estimates for "pressure" interpolants.

$$\|\pi_h^V w - \widetilde{\pi}_h^V w\| \le C_V \, E(\mathfrak{T}_h, \widetilde{\mathcal{T}}_h, w), \tag{217}$$

where the constant C_V does not depend on h and the quantity E is defined by the relations

$$E(\mathfrak{T}_h, \widetilde{\mathcal{T}}_h, w) = \max_{i=1,2,\dots N} E_i(w),$$

$$E_i(w) = \max_{j=1,\dots m_i} \left| [w]_{T_j^i} - [w]_{E^i} \right| .$$

Indeed, take a cell $\mathbf{E^i}$ and a simplex in it.

$$\int_{T_j^i} (\pi_h^V w - \widetilde{\pi}_h^V w)^2 \, dx = |T_j^i| \left(\frac{1}{|E^i|} \int_{E^i} w \, dx - \frac{1}{|T_j^i|} \int_{T_j^i} w \, dx \right)^2 =$$

$$= |T_j^i| \left(\frac{|T_j^i| \int_{E^i} w \, dx - |E^i| \int_{T_j^i} w \, dx}{|E^i||T_j^i|} \right)^2 =$$

$$= |T_j^i| \left(\frac{|T_j^i| \int_{E^i \setminus T_j^i} w \, dx - |E^i \setminus T_j^i| \int_{T_j^i} w \, dx}{|E^i||T_j^i|} \right)^2 =$$

$$= \frac{|T_j^i||E^i \setminus T_j^i|^2}{|E^i|^2} \left([w]_{E^i \setminus T_j^i} - [w]_{T_j^i} \right)^2 .$$

Now, we apply the estimate

$$\left([\mathbf{w}]_{\mathbf{E^i}\setminus\mathbf{T^i_j}} - [\mathbf{w}]_{\mathbf{T^i_j}}\right)^2 \leq \mathbf{2}\left[\left([\mathbf{w}]_{\mathbf{E^i}\setminus\mathbf{T^i_j}} - [\mathbf{w}]_{\mathbf{E^i}}\right)^2 + \left([\mathbf{w}]_{\mathbf{T^i_j}} - [\mathbf{w}]_{\mathbf{E^i}}\right)^2\right]$$

and note that

$$[\mathbf{w}]_{\mathbf{E^i}\setminus\mathbf{T^i_j}} - [\mathbf{w}]_{\mathbf{E^i}} \leq \mathbf{E_i(w)}.$$

By these arguments we obtain

$$\int_{\mathbf{T^i_j}} (\pi^{\mathbf{V}}_{\mathbf{h}}\mathbf{w} - \widetilde{\pi}^{\mathbf{V}}_{\mathbf{h}}\mathbf{w})^2 \,\mathbf{dx} \leq \mathbf{2}|\mathbf{T^i_j}|\mu_{\mathbf{i}}^2\mathbf{E}_{\mathbf{i}}^2\mathbf{(w)}.$$

By summing over all simplexes, than over all cells we obtain (279). Since $\mu_i \leq 1$, $\mathbf{C_V = 2}$.

Similar estimate holds true for the interpolants of fluxes

$$\|\pi_{\mathbf{h}}^{\mathbf{Q}}\mathbf{q} - \widetilde{\pi}_{\mathbf{h}}^{\mathbf{Q}}\mathbf{q}\|_{\mathbf{div},\Omega} \leq \bar{\sigma}\mathbf{E}(\mathfrak{T}_{\mathbf{h}}, \widetilde{\mathcal{T}}_{\mathbf{h}}, \mathbf{divq}), \tag{218}$$

Rate convergence estimates for p_h and u_h

By the above results we arrive at *a priori rate convergence estimates* for approximations on distorted meshes (see Yu. Kuznetsov and S. Repin, JNM, 2005).

$$\|\mathbf{p} - \mathbf{p_h}\|_{\mathbf{div},\Omega} + \|\mathbf{u} - \mathbf{u_h}\| \leq$$
$$\leq \mathbf{C}\Big[\mathbf{h}(\|\mathbf{p}\|_{1,\Omega} + \|\mathbf{divp}\|_{1,\Omega} + \|\nabla \mathbf{u}\|) +$$
$$+\mathbf{E}(\mathfrak{T}_\mathbf{h}, \widetilde{\mathcal{T}}_\mathbf{h}, \mathbf{divp}) + \mathbf{E}(\mathfrak{T}_\mathbf{h}, \widetilde{\mathcal{T}}_\mathbf{h}, \mathbf{u})\Big],$$

A posteriori estimates for cell approximations

$$\|\mathbf{p} - \mathbf{p_h}\| \leq \sqrt{2}\|\nabla \mathbf{u_h} - \widehat{\mathbf{p}}_\mathbf{h}\| + 2\mathbf{C}_\Omega \|\mathbf{div}\mathbf{p_h} + \mathbf{f}\|\,,$$

Since $-\mathbf{div}\mathbf{p_h} = [\mathbf{f}]_\mathbf{E}$, we can rewrite this upper bound as

$$\|\mathbf{p} - \mathbf{p_h}\| \leq \sqrt{2}\|\nabla \mathbf{u_h} - \widehat{\mathbf{p}}_\mathbf{h}\| + 2\mathbf{C}_\Omega \|[\mathbf{f}]_\mathbf{E} - \mathbf{f}\|\,,$$

Lecture 8

A POSTERIORI ESTIMATES FOR PROBLEMS IN THE THEORY OF VISCOUS INCOMPRESSIBLE FLUIDS

Lecture plan

- Mathematical models of viscous fluids;
- Stokes problem;
- Inf-sup condition ;
- A posteriori estimates for solenoidal approximations ;
- A posteriori estimates for non-solenoidal approximations;
- A posteriori estimates for problems with condition divv $= \phi$;
- A posteriori estimates for problems on a subspace.
- Bingham fluids. A posteriori estimates.

Coordinates of particles at $t = 0$ are denoted $\mathbf{X}$ and called Lagrangian coordinates. They serve as particles labels. Then, the trajectory of an individual particle is given by the relation

$$\mathbf{x} = \mathbf{x}(\mathbf{X}, \mathbf{t}),$$

where $\mathbf{x}$ denote the Cartesian coordinates of particles at the moment t. They are called Euler coordinates.
Motion equations describing evolution of a media have the form

$$\rho_0 \frac{\partial \mathbf{v}}{\partial \mathbf{t}} (\mathbf{x}, \mathbf{t}) + \mathbf{v_i} (\mathbf{x}, \mathbf{t}) \frac{\partial \mathbf{v}}{\partial \mathbf{x_i}} (\mathbf{x}, \mathbf{t}) = \mathbf{div}\ \boldsymbol{\sigma} (\mathbf{x}, \mathbf{t}) + \mathbf{f} (\mathbf{x}, \mathbf{t}) \qquad (219)$$

Here $\boldsymbol{\sigma}$ is the effective stress,
$\mathbf{v_i} \frac{\partial \mathbf{v}}{\partial \mathbf{x_i}}$ is the so–called convective term which is often presented as

$$(\mathbf{v} \cdot \nabla) \mathbf{v} (\mathbf{x}, \mathbf{t})$$

and

$$\varepsilon (\mathbf{v}) = \frac{1}{2} (\mathbf{v_{i,j}} + \mathbf{v_{j,i}}) \tag{220}$$

is the tensor of small strains.
In the majority of models liquids are assumed to be incompressible, what means that

$$\mathbf{div v} = \mathbf{0}.$$

In view of this fact

$$\varepsilon (\mathbf{v}) = \varepsilon^D (\mathbf{v}).$$

Constitutive relation are usually given in the form

$$\boldsymbol{\sigma} = -\mathbf{p}\mathbb{I} + \boldsymbol{\tau}, \tag{221}$$

where $\mathbf{p}$ is the pressure and $\mathbb{I}$ is the unit element of $\mathbb{M}_s^{n\times n}$ and "deviatoric stress" is defined by the relation

$$\boldsymbol{\tau} \in \partial \mathbf{W}(\boldsymbol{\varepsilon}) \tag{222}$$

where $\mathbf{W} : \mathbb{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}} \longmapsto \mathbb{R}_+$ is the dissipative potential and $\partial\mathbf{W}$ stands for subdifferential.

Boundary conditions

Two main types of the problems:
Flow of a fluid in a container (basin).

$$\Omega_0 = \Omega_t = \Omega, \qquad t \geq 0; \qquad \mathbf{v} = \mathbf{u_0} \quad \partial\boldsymbol{\Omega} \tag{223}$$

$\mathbf{v} = \mathbf{0}$ on $\partial\Omega$ is called the "adhesion" or "no–slip" condition.
Flow in an "open" domain.
In such a case, on a part of the boundary we set

$$\boldsymbol{\sigma}(\mathbf{x}, \mathbf{t})\, \boldsymbol{\nu}(\mathbf{x}, \mathbf{t}) = \mathbf{F}(\mathbf{x}, \mathbf{t}), \qquad \mathbf{x} \in \partial\boldsymbol{\Omega}_\mathbf{t}, \qquad \mathbf{t} \geq \mathbf{0} \tag{224}$$

Initial condition.

$$\mathbf{v}(\mathbf{x}, \mathbf{0}) = \varphi(\mathbf{x}) \qquad \mathbf{x} \in \boldsymbol{\Omega}_\mathbf{0} \tag{225}$$

Now the task is to find $\mathbf{v}$, $\mathbf{p}$ $\boldsymbol{\tau}$ such that

$$\rho_0 \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v_i} \frac{\partial \mathbf{v}}{\partial x_i} - \mathbf{div}\ \boldsymbol{\sigma} = \mathbf{f} \tag{226}$$

$$\mathbf{div}\ \mathbf{v} = \mathbf{0} \qquad (\mathbf{x}, \mathbf{t}) \in \mathbf{Q} := \boldsymbol{\Omega} \times (\mathbf{0}, +\infty) \tag{227}$$

$$\boldsymbol{\sigma} = -\mathbf{p}\mathbb{I} + \boldsymbol{\tau}, \qquad \boldsymbol{\tau} \in \partial \mathbf{W}(\boldsymbol{\varepsilon}) \tag{228}$$

$$\mathbf{v}(\mathbf{x}, \mathbf{t}) = \mathbf{0} \qquad (\mathbf{x}, \mathbf{t}) \in \partial_1 \boldsymbol{\Omega} \times (\mathbf{0}, +\infty) \tag{229}$$

$$\boldsymbol{\sigma\nu} = \mathbf{F} \qquad (\mathbf{x}, \mathbf{t}) \in \partial_2 \boldsymbol{\Omega} \times (\mathbf{0}, +\infty) \tag{230}$$

$$\mathbf{v}(\mathbf{x}.\mathbf{0}) = \varphi(\mathbf{x}) \qquad \mathbf{x} \in \boldsymbol{\Omega}$$

Bingham type fluids

$$\mathbf{W}(\varepsilon) = \frac{\mu}{\mathbf{m}} |\varepsilon|^{\mathbf{m}} + \mathbf{k}_* |\varepsilon| \tag{231}$$

A particular case with $\mathbf{k}_* = \mathbf{0}$ and $\mathbf{m} = \mathbf{2}$ leads to

$$\boldsymbol{\sigma} = -\mathbf{p}\mathbb{I} + \frac{\partial \mathbf{W}}{\partial \boldsymbol{\varepsilon}} = -\mathbf{p}\mathbb{I} + \mu\boldsymbol{\varepsilon},$$

Then

$$\mathbf{div}\ \boldsymbol{\sigma} = -\nabla \mathbf{p} + \mu \triangle \mathbf{v}$$

and we arrive at the Navier–Stokes system

$$\begin{cases} \rho_0 \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v_i} \frac{\partial \mathbf{v}}{\partial \mathbf{x_i}} - \mu \triangle \mathbf{v} = \mathbf{f} + \nabla \mathbf{p} \\ \mathbf{div}\ \mathbf{v} = \mathbf{0} \end{cases} \tag{232}$$

If $k_* > 0$ then it is typical that the stagnant zones with $\varepsilon(\mathbf{v}) \equiv 0$ arise.

Power law models

$$\mathbf{W}(\varepsilon) = \mu_\infty |\varepsilon|^2 + \mu_0 \left(\mathbf{1} + |\varepsilon|^2\right)^{\frac{\mathbf{p}}{2}} \tag{233}$$

$$\mathbf{W}(\varepsilon) = \mu_\infty |\varepsilon|^2 + \mu_0 |\varepsilon|^{\mathbf{p}} \tag{234}$$

Here $\mathbf{p} \in]\mathbf{1}, +\infty[$, $\mu_\infty \geq 0$ $\mu_0 \geq 0, \mu_\infty \mu_0 \neq 0$.

Powel–Euring models

$$\mathbf{W}(\varepsilon) = \mu_\infty |\varepsilon|^2 + \mu_1 |\varepsilon| \ln(\mathbf{1} + |\varepsilon|) \tag{235}$$

General system

In all the cases we have the system

$$\begin{cases} \rho_0 \frac{\partial \mathbf{v}}{\partial t} + \mathbf{v_i} \frac{\partial \mathbf{v}}{\partial x_i} - \frac{\partial W}{\partial \varepsilon}(\varepsilon(\mathbf{v})) = \mathbf{f} + \nabla \mathbf{p} \\ \mathbf{div}\ \mathbf{v} = 0 \end{cases} \tag{236}$$

In case of slow motions the term $\mathbf{v} \cdot (\nabla \mathbf{v})$ is usually neglected. Then, we arrive at simplified models among which the most known is the Stokes model.

Spaces of solenoidal functions

$$\dot{\mathbf{J}}^{\infty}(\Omega) = \{\mathbf{u} \in \mathbf{C}_0^{\infty}(\Omega), \mathbf{div u} = \mathbf{0}, \mathrm{supp}\mathbf{u} \subset \Omega\},$$

$$\overset{\circ}{\mathbf{J}}(\Omega) := \text{ closure of } \dot{\mathbf{J}}^{\infty} \text{ in the topology of } \mathbf{L}_2(\Omega, \mathbb{R}^{\mathbf{d}}),$$

$$\overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\Omega)(\Omega) := \text{ closure of } \dot{\mathbf{J}}^{\infty} \text{ in the topology of } \mathbf{H}^{\mathbf{1}}(\Omega, \mathbb{R}^{\mathbf{d}}).$$

Navier–Stokes equation

At present Navier–Stokes problem dominates among the models describing the behavior of viscous incompressible fluids. It is to find $\mathbf{u}(\mathbf{x},\mathbf{t}) \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega})$ and $\mathbf{p}(\mathbf{x},\mathbf{t}) \in \overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\mathbf{\Omega})$ such that

$$
\begin{aligned}
&\mathbf{u_t} - \nu\mathbf{\Delta u} + \mathbf{div}(\mathbf{u}\otimes\mathbf{u}) = \mathbf{f} - \nabla\mathbf{p} \qquad \text{in } \mathbf{\Omega},,\\
&\mathbf{u}(\mathbf{x},\mathbf{0}) = \varphi(\mathbf{x}),\\
&\mathbf{u} = \mathbf{u_0} \quad \text{on } \mathbf{\Gamma_D}\\
&\varepsilon(\mathbf{u})\cdot\boldsymbol{\nu} + \mathbf{p}\boldsymbol{\nu} = \mathbf{g_N} \quad \text{on}\mathbf{\Gamma_N}.
\end{aligned}
$$

From the mathematical point of view NS is still a mystery. Existence of a unique solution in 3D is not yet proved even for $(0, T] \times \mathbb{R}^n$.

It is known that for sufficiently regular solenoidal $\phi(\mathbf{x})$ there exists a weak Leray-Hopf solution, i.e., a function

$$\mathbf{u} \in \mathbf{L}^{\infty}(\mathbf{0}, \mathbf{T}; \mathbf{L}^{\mathbf{2}}(\mathbb{R}^{\mathbf{n}})) \cap \mathbf{L}^{\mathbf{2}}(\mathbf{0}, \mathbf{T}; \mathbf{H}^{\mathbf{1}}(\mathbb{R}^{\mathbf{n}}))$$

Proving (or presenting a contr–example) of that NS equation possesses in $(0, T] \times \mathbb{R}^n$ a smooth solution provided that initial data are sufficiently regular forms one of the
Millennium Prize Problems
stated by the Clay Mathematical Institute.

From the introduction to the Third Millennium Prize Problem:
"...Although these (NS) equations were written down in the 19th Century, our understanding of them remains minimal. The challenge is to make substantial progress toward a mathematical theory which will unlock the secrets hidden in the Navier-Stokes equations."

However, discrete (semidiscrete) analogs of NS equation are actively used in the Mathematical Modeling.

For example:

$$\frac{\mathbf{u}^k - \mathbf{u}^{k-1}}{\Delta t} - \nu \Delta \mathbf{u}^k + \mathbf{div}(\mathbf{u}^{k-1} \otimes \mathbf{u}^{k-1}) = \mathbf{f} - \nabla \mathbf{p}^k \qquad \text{in } \Omega,$$
$$\mathbf{div} \mathbf{u}^k = 0$$

and

$$\frac{\mathbf{u}^k - \mathbf{u}^{k-1}}{\Delta t} - \nu \Delta \mathbf{u}^k + \mathbf{div}(\mathbf{u}^{k-1} \otimes \mathbf{u}^{k}) = \mathbf{f} - \nabla \mathbf{p}^k \qquad \text{in } \Omega,$$
$$\mathbf{div} \mathbf{u}^k = 0.$$

For these reasons, an intensive investigations has been devoted to *linearizations* of NS equations, in particulat to Stokes

$$
\begin{aligned}
&\mathbf{u_t} - \nu\mathbf{\Delta u} = \mathbf{f} - \nabla\mathbf{p} \qquad \text{in } \mathbf{\Omega},\\
&\mathbf{u}(\mathbf{x},\mathbf{0}) = \varphi(\mathbf{x}),\\
&\mathbf{u} = \mathbf{u_0} \quad \text{on } \mathbf{\Gamma_D}\\
&\varepsilon(\mathbf{u})\cdot\nu + \mathbf{p}\nu = \mathbf{g_N} \quad \text{on}\mathbf{\Gamma_N}
\end{aligned}
$$

and Oseen

$$
\begin{aligned}
&\mathbf{u_t} - \nu\mathbf{\Delta u} + \mathbf{div}(\mathbf{a}\otimes\mathbf{u}) = \mathbf{f} - \nabla\mathbf{p} \qquad \text{in } \mathbf{\Omega},\\
&\mathbf{u}(\mathbf{x},\mathbf{0}) = \varphi(\mathbf{x}), \qquad \mathbf{diva} = \mathbf{0},\\
&\mathbf{u} = \mathbf{u_0} \quad \text{on } \mathbf{\Gamma_D}\\
&\varepsilon(\mathbf{u})\cdot\nu + \mathbf{p}\nu = \mathbf{g_N} \quad \text{on}\mathbf{\Gamma_N}
\end{aligned}
$$

problems.

NS equations with rotation

In certain cases (e.g., ocean modeling) NS equations should take into account Earth rotation. Then, an additional term arises and the equation comes in the form

$$\mathbf{u_t} - \nu \mathbf{\Delta u} + (\mathbf{u} \cdot \nabla \mathbf{u}) + \omega \times \mathbf{u} = \mathbf{f} - \nabla \mathbf{p} \qquad \text{in } \mathbf{\Omega},$$

where

$$\omega = |\omega| \mathbf{e_3}.$$

is the parameter of the rotation intensity.

Computer simulation methods

A significant part of the difficulties arising in the process of solving such problems is related to the **incompressibility condition divu = 0**.
Typically, this condition is taken into account by projecting of a discrete solution to the set of solenoidal fields or by introducing appropriate penalty terms. A detailed exposition of the numerical methods can be found, e.g., in the works of (see the List of Literature) J. Chen, A. Chorin, W. E and J. G. Liu, M. Feistauer, M. Ganzburger, V. Girault, G. Heywood, R. Rannacher, P. A. Raviart, R. Temam.
Stationary problems are often solved by passing to a minimax formulation and using the so–called mixed approximations for the velocity and pressure fields (see, e.g., F. Brezzi and J. Duglas, F. Brezzi and M. Fortin).

Linear models in the theory of fluids

Classical formulation of the Stokes problem: find a vector–valued function $\mathbf{u}$ (velocity) and a scalar–valued function $\mathbf{p}$ (pressure) that satisfy the relations

$$-\nu\boldsymbol{\Delta}\mathbf{u} = \mathbf{f} - \nabla\mathbf{p} \qquad \text{in } \Omega, \tag{237}$$

$$\mathbf{div}\,\mathbf{u} = \mathbf{0} \qquad \text{in } \Omega, \tag{238}$$

$$\mathbf{u} = \mathbf{u_0} \qquad \text{on } \partial\Omega, \tag{239}$$

where $\mathbf{u_0}$ is a given function such that $\mathbf{div}\,\mathbf{u_0} = \mathbf{0}$.

Here Δ denotes the Laplacian of a vector field:

$$\mathbf{\Delta u} = (\mathbf{\Delta u_i})\,\mathbf{e_i},$$

e_i are the cartesian unit vectors.
Multiply (237) by a function $\mathbf{v} \in \dot{\mathbf{J}}^\infty$. Then,

$$\int_\Omega \nabla \mathbf{p} \cdot \mathbf{v}\,\mathbf{dx} = \int_\Omega \mathbf{p}\ \mathbf{div v}\,\mathbf{dx} = \mathbf{0},$$

and we arrive at the relation

$$\nu((\mathbf{u},\mathbf{v})) = (\mathbf{f},\mathbf{v}) \quad \forall \mathbf{v} \in \dot{\mathbf{J}}^\infty(\mathbf{\Omega}), \tag{240}$$

where $((\mathbf{u},\mathbf{v})) = \int_\Omega \varepsilon(\mathbf{u}) : \varepsilon(\mathbf{v})\mathbf{dx}$. Above identity can be extended by continuity to the whole $\mathring{\mathbf{J}}^1_2(\mathbf{\Omega})$.

Weak formulation of the Stokes problem

. Find $\mathbf{u} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega})$, such that $\mathbf{u} = \mathbf{0}$ on$\partial\Omega$ and

$$\mu((\mathbf{u}, \mathbf{v})) = (\mathbf{f}, \mathbf{v}) \quad \forall \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega})(\mathbf{\Omega}), \tag{241}$$

This formulation is related to the energy functional

$$\mathbf{I}(\mathbf{v}) = \mu \int_{\mathbf{\Omega}} |\nabla \mathbf{v}|^{\mathbf{2}} \, \mathbf{dx} - \mathbf{2} \int_{\mathbf{\Omega}} \mathbf{f} \, \mathbf{v} \, \mathbf{dx} \tag{242}$$

where $|\nabla \mathbf{v}|^{\mathbf{2}} = |\nabla \mathbf{v_1}|^{\mathbf{2}} + ... + |\nabla \mathbf{v_n}|^{\mathbf{2}}$.

Theorem

Generalized solution of the Stokes problem minimizes the functional $\mathbf{I}$ *over* $\overset{\circ}{\mathbf{J}}{}_2^1(\Omega)$.

Proof. Assume that

$$\mu((\mathbf{u},\mathbf{v})) = (\mathbf{f},\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{V}.$$

Since $\|\mathbf{u}-\mathbf{v}\|^2 \geq \mathbf{0}$, we have

$$\mu\|\mathbf{u}\|^2 + \mu\|\mathbf{v}\|^2 - 2\mu((\mathbf{u},\mathbf{v})) \geq \mathbf{0}. \tag{243}$$

On the other hand,

$$\mathbf{I}(\mathbf{u}) = \mu\|\mathbf{u}\|^2 - 2(\mathbf{f},\mathbf{u}) = \mu\|\mathbf{u}\|^2 - 2\mu((\mathbf{u},\mathbf{u})) = -\mu\|\mathbf{u}\|^2$$

Therefore, we rewrite (243) as follows:

$$\mu\|\mathbf{v}\|^2 - 2\mu((\mathbf{u},\mathbf{v})) \geq \mathbf{I}(\mathbf{u}),$$

Therefore,

$$\mathsf{I}(\mathbf{v}) \geq \mathsf{I}(\mathbf{u}) \quad \forall \mathbf{v} \in \mathbf{V}. \tag{244}$$

Assume that $\mathbf{u}$ is the minimizer, then for any $\mathbf{v} \in \mathbf{V}$

$$\begin{gathered} \mathsf{I}(\mathbf{u} + \lambda \mathbf{v}) \geq \mathsf{I}(\mathbf{u}), \\ \mu \left\|\mathbf{u}\right\|^2 + 2\lambda\mu((\mathbf{u}, \mathbf{v})) + \mu\lambda^2 \left\|\mathbf{v}\right\|^2 - 2(\mathbf{f}, \mathbf{u} + \lambda \mathbf{v}) \geq \\ \geq \mu \left\|\mathbf{u}\right\|^2 - 2(\mathbf{f}, \mathbf{u}), \end{gathered}$$

and, consequently,

$$\mu((\mathbf{u}, \mathbf{v})) - (\mathbf{f}, \mathbf{v}) \geq -\frac{\lambda}{2} \left\|\mathbf{v}\right\|^2 \implies \mu((\mathbf{u}, \mathbf{v})) - (\mathbf{f}, \mathbf{v}) = 0 \quad \forall \mathbf{v} \in \mathbf{V}.$$

Nomenclature

Next, $\mathbf{W} := \mathbf{W_2^1}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}})$ and $\mathbf{\Sigma} := \mathbf{L_2}(\mathbf{\Omega}, \mathbb{M}^{\mathbf{d}\times\mathbf{d}})$, where $\mathbb{M}^{d\times d}$ is the space of symmetric $\mathbf{d}\times\mathbf{d}$ matrixes (tensors), whose scalar product is denoted by two dots.

$\mathbf{W_0}$ is a subspace of $\mathbf{W}$ that contains functions with zero traces on $\partial\Omega$.

$\mathbf{W_0} + \mathbf{u_0}$ contains functions of the form $\mathbf{w} + \mathbf{u_0}$, where $\mathbf{w} \in \mathbf{V_0}$.

Analogously, $\overset{\circ}{\mathbf{J}}{}_{\mathbf{2}}^{\mathbf{1}}(\mathbf{\Omega}) + \mathbf{u_0}$ contains functions of the form $\mathbf{w} + \mathbf{u_0}, \mathbf{w} \in \overset{\circ}{\mathbf{J}}{}_{\mathbf{2}}^{\mathbf{1}}(\mathbf{\Omega})$.

The operator $\boldsymbol{\varepsilon}(\mathbf{v}) := \frac{\mathbf{1}}{\mathbf{2}}(\nabla\mathbf{v} + (\nabla\mathbf{v})^{\mathbf{T}})$ acts from $\mathbf{W}$ to $\mathbf{\Sigma}$.

We will also use the Hilbert space $\mathbf{\Sigma}_{\mathrm{div}}(\mathbf{\Omega})$, which is a subspace of $\mathbf{\Sigma}$ that contains tensor–valued functions $\boldsymbol{\tau}$, such that $\mathbf{div}\boldsymbol{\tau} \in \mathbf{L_2}$. The scalar product in this space is defined by the relation

$$(\boldsymbol{\tau}, \boldsymbol{\eta}) := \int_{\mathbf{\Omega}} (\boldsymbol{\tau} : \boldsymbol{\eta} + \mathbf{div}\boldsymbol{\tau} \cdot \mathbf{div}\boldsymbol{\eta})\, \mathbf{dx}\,.$$

As before, $\mathring{\mathbf{L}}_{\mathbf{2}}(\mathbf{\Omega})(\mathbf{\Omega})$ denotes the space of square summable functions with zero mean. Henceforth, we assume that

$$\mathbf{f} \in \mathbf{L_2}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}}), \quad \mathbf{u_0} \in \mathbf{W_2^1}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}}),$$

Generalized solution can be defined by the **integral identity**. It is a function $\mathbf{u} \in \overset{\circ}{\mathbf{J}}{}_2^1(\Omega) + \mathbf{u_0}$ that meets the relation

$$\int_\Omega \nu \nabla(\mathbf{u}) : \nabla(\mathbf{v})\, \mathbf{dx} = \int_\Omega \mathbf{f} \cdot \mathbf{v}\, \mathbf{dx} \quad \forall \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}_2^1(\Omega). \tag{245}$$

It is well known that $\mathbf{u}$ exists and unique and can be viewed as the minimizer of the functional

$$\mathbf{I}(\mathbf{v}) = \int_\Omega \left(\frac{\nu}{2} |\nabla(\mathbf{v})|^2 - \mathbf{f} \cdot \mathbf{v} \right) \mathbf{dx}$$

on the set $\overset{\circ}{\mathbf{J}}{}_2^1(\Omega) + \mathbf{u_0}$. Thus, the problem

$$\inf_{\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}_2^1(\Omega) + \mathbf{u_0}} \mathbf{I}(\mathbf{v})$$

presents a **variational formulation** of the Stokes problem.

Existence of a minimizer follows from known properties of convex lower semicontinuous functionals.
In addition, the Stokes problem can be presented in a **minimax form**.
Let $\mathbf{L} : (\mathbf{W_0} + \mathbf{u_0}) \times \overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\mathbf{\Omega}) \to \mathbb{R}$ be defined as follows:

$$\mathbf{L(v,q)} = \int_{\mathbf{\Omega}} \left(\frac{\boldsymbol{\nu}}{\mathbf{2}} |\nabla \mathbf{v}|^{\mathbf{2}} - \mathbf{f} \cdot \mathbf{v} - \mathbf{q div v} \right) \mathbf{dx}.$$

Now, $\mathbf{u}$ and $\mathbf{p}$ are defined as a saddle–point that satisfies the relations

$$\mathbf{L(u,q)} \leq \mathbf{L(u,p)} \leq \mathbf{L(v,p)} \qquad \forall \mathbf{v} \in \mathbf{W_0} + \mathbf{u_0},\ \mathbf{p} \in \overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\mathbf{\Omega}).$$

Extension of solenoidal fields and related results

First, we recall some basic results that has been established when the **solvability of the Stokes** problem was investigated. Works of O.A. Ladyzhenskaya have made a considerable and widely acknowledged contribution to the mathematical theory of viscous incompressible fluids.

The first principal result states that a solenoidal field can be extended inside a domain such that the norm of the extended field is subject to the norm of the boundary trace (see O.A. Ladyzhenskaya *Mathematical problems in the dynamics of a viscous incompressible fluid*. Nauka, Moscow, 1970 and
O.A Ladyzhenskaya and V.A. Solonnikov Some problems of vector analysis, and generalized formulations of boundary value problems for the Navier-Stokes equation, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 59(1976), 81–116, 256).

Lemma 1.

For any vector–valued function $\mathbf{a} \in \mathbf{W}_2^{1/2}(\partial\Omega)$ satisfying the condition $\int_{\partial\Omega} \mathbf{a} \cdot \boldsymbol{\nu}\, \mathbf{dx} = \mathbf{0}$ there exists a function $\bar{\mathbf{u}} \in \mathbf{W_0}$ such that $\mathbf{div}\bar{\mathbf{u}} = 0$ and

$$\|\nabla \bar{\mathbf{u}}\| \leq \boldsymbol{\kappa_1}(\Omega)\|\mathbf{a}\|_{1/2,\partial\Omega}, \tag{246}$$

where $\boldsymbol{\kappa}_1(\Omega)$ is a positive constant that depends on Ω.

This lemma implies another proposition, which is of grate importance for the analysis of problems defined on solenoidal fields.

Lemma 2

For any $\mathbf{f} \in \mathring{\mathbf{L}}_2(\Omega)$ there exists a function $\bar{\mathbf{u}} \in \mathbf{W}_0$ satisfying the relation $\mathbf{div}\bar{\mathbf{u}} = \mathbf{f}$ and the condition

$$\|\nabla \bar{\mathbf{u}}\| \leq \boldsymbol{\kappa}_2(\Omega)\|\mathbf{f}\|, \tag{247}$$

where $\boldsymbol{\kappa}_2(\Omega)$ is a positive constant that depends on Ω.

Lemma 2 implies several important corollaries that we discuss below.

Inf-Sup condition

Lemma 2 is related to the inequality known in the literature as the **Inf-Sup**– or **LBB** (**Ladyzhenskaya–Babuška–Brezzi)–condition that reads: there exists a positive constant C such that**

$$\inf_{\substack{\phi \in \overset{\circ}{L}_2(\Omega) \\ \phi \neq 0}} \sup_{\substack{\mathbf{w} \in \mathbf{W_0} \\ \mathbf{w} \neq \mathbf{0}}} \frac{\int_\Omega \phi \, \mathbf{div w} \, \mathbf{dx}}{\|\phi\| \, \|\nabla \mathbf{w}\|} \geq \mathbf{C} \, . \tag{248}$$

Inf-Sup condition (248) was established in the papers by I. Babuška The finite element method with Lagrangian multipliers, *Numer. Math.*, 20(1973) and F. Brezzi, On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers, *R.A.I.R.O., Annal. Numer.*, 8 (1974). They used its discrete analogs for proving the convergence of finite–dimensional approximations in various problems related to the theory of viscous incompressible fluids.

Lemma 2 implies LBB condition

By Lemma 2, any $\phi \in \mathring{\mathbf{L}}_2(\Omega)$ has a counterpart function $\mathbf{v}_\phi \in \mathbf{W_0}$ that meets the conditions

$$\mathbf{div}\mathbf{v}_\phi = \phi, \quad \|\nabla \mathbf{v}_\phi\| \le \boldsymbol{\kappa}_2(\Omega)\|\phi\|.$$

In this case,

$$\sup_{\mathbf{v}\in\mathbf{W_0},\mathbf{w}\neq\mathbf{0}} \frac{\int_\Omega \phi \mathbf{div}\mathbf{v}\,\mathbf{dx}}{\|\nabla \mathbf{v}\|\,\|\phi\|} \ge \frac{\int_\Omega \phi \mathbf{div}\mathbf{v}_\phi\,\mathbf{dx}}{\|\nabla \mathbf{v}_\phi\|\,\|\phi\|} = \frac{\|\phi\|}{\|\nabla \mathbf{v}_\phi\|} \ge \frac{\mathbf{1}}{\boldsymbol{\kappa}_2(\Omega)}$$

and, consequently, Inf-Sup condition holds with

$$\boxed{\mathbf{C} = \frac{1}{\boldsymbol{\kappa}_2(\Omega)}.}$$

It is easy to observe that the Inf-Sup condition can be presented in the form

$$\sup_{\substack{\mathbf{w}\in\mathbf{W_0}\\ \mathbf{w}\neq\mathbf{0}}} \frac{\int_\Omega \mathbf{p}\,\mathrm{div}\mathbf{w}\,dx}{\|\nabla\mathbf{w}\|} \geq \mathbf{C}\,\|\mathbf{p}\| \quad \text{for all } \mathbf{p}\in\mathring{\mathbf{L}}_2(\Omega).$$

We may consider the expression in the left–hand side of the above inequality as the norm of $\nabla\mathbf{p}$ in the space topologically dual to $\mathbf{W_0}$, namely

$$\mathbf{[}\nabla\mathbf{p}\mathbf{]} := \sup_{\mathbf{w}\in\mathbf{W_0}} \frac{<\nabla\mathbf{p},\mathbf{w}>}{\|\nabla\mathbf{w}\|}.$$

Then, we arrive to the Nečas inequality.

Nečas inequality

$$\|\mathbf{p}\| \leq \boldsymbol{\kappa}_2 \, \mathbf{I} \nabla \mathbf{p} \mathbf{I} \qquad \forall \ \mathbf{p} \in \overset{\circ}{\mathbf{L}}_2(\Omega) \,, \tag{249}$$

A simple proof of the Nečas inequality for domains with Lipschitz boundaries can be found in the paper by
J. Bramble. *A proof of the inf-sup condition for the Stokes equations on Lipschitz domains*, Math. Models Methods Appl. Sci. **13** (2003), no. 3, 361–371.
In the later paper, it is also shown that the well–known Korn's inequality follows from Inf-Sup condition.
Constants $\mathbf{C}$ and $\boldsymbol{\kappa}_2$ play an important role in the numerical analysis of the Stokes problem as well as in the theoretical one.

Existence of a saddle point

Existence of a saddle point of $\mathbf{L}(\mathbf{v}, \mathbf{q})$ follows from Lemma 2 and known results of the minimax theory. In a simplified version these results reads:

Lagrangian $\mathbf{L}(\mathbf{v}, \mathbf{q})$ possess a saddle point provided that
(a) it is convex and continuous with respect to the first variable and concave and continuous with respect to the second one;

(b) for a certain $\bar{\mathbf{q}}$ the functional $\mathbf{v} \mapsto \mathbf{L}(\mathbf{v}, \bar{\mathbf{q}})$ is coercive (or the set of admissible $\mathbf{v}$ is compact);

(c) or a certain $\bar{\mathbf{v}}$ the functional $\mathbf{q} \mapsto -\mathbf{L}(\bar{\mathbf{v}}, \mathbf{q})$ is coercive (or the set of admissible $\mathbf{q}$ is compact.)

Since

$$\mathbf{J}(\mathbf{v}) = \sup_{\mathbf{q}\in\mathbf{\Sigma}} \mathbf{L}(\mathbf{v},\mathbf{q}) \geq \mathbf{L}(\bar{\mathbf{q}},\mathbf{v}),$$

we observe that (b) means that $\mathbf{J}(\mathbf{v})$ is coercive. Analogously, (c) means that the functional $-\mathbf{I}(\mathbf{q})$, where

$$\mathbf{I}(\mathbf{q}) = \inf_{\mathbf{q}\in\mathbf{V_0}+\mathbf{u_0}} \mathbf{L}(\mathbf{v},\mathbf{q}) \leq \mathbf{L}(\mathbf{q},\bar{\mathbf{v}}),$$

is coercive.
In other words, for a continuous convex-concave Lagrangian existence of a saddle point mainly depends on the coercivity properties of the two dual functionals generated by it.

Let us apply these results to the Stokes problem. It is easy to see that for any $\mathbf{q} \in \mathring{\mathbf{L}}_2(\Omega)$ the mapping

$$\mathbf{v} \mapsto \mathbf{L}(\mathbf{v}, \mathbf{q}) = \int_\Omega \left(\frac{\nu}{2} |\nabla \mathbf{v}|^2 - \mathbf{f} \cdot \mathbf{v} - \mathbf{q} \mathrm{div} \mathbf{v} \right) \mathbf{dx}.$$

is convex and continuous (in $\mathbf{W}$) and there exists am element $\bar{\mathbf{q}} \in \mathring{L}_2(\Omega)$ (e.g., $\bar{\mathbf{q}} = 0$) such that $\mathbf{L}(\mathbf{v}, \bar{\mathbf{q}}) \to +\infty$ if $\|\mathbf{v}\|_{\mathbf{V}} \to +\infty$. The mapping $\mathbf{q} \mapsto \mathbf{L}(\mathbf{v}, \mathbf{q})$ is affine and continuous (in $\mathring{\mathbf{L}}_2(\Omega)$) for any $\mathbf{v} \in \mathbf{V}$. Therefore, existence of a saddle point is guaranteed provided that the coercivity condition

$$\lim_{\|\mathbf{q}\| \to +\infty} \inf_{\mathbf{v} \in \mathbf{W_0} + \mathbf{u_0}} \mathbf{L}(\mathbf{v}, \mathbf{q}) = -\infty \tag{250}$$

is established. By Lemma 2 we can prove this fact.

Consider the functional

$$\mathsf{I}(\mathbf{q}) := \inf_{\mathbf{v}\in \mathbf{W_0}+\mathbf{u_0}} \mathsf{L}(\mathbf{v},\mathbf{q})$$

and the variational problem

$$\mathsf{I}(\mathbf{p}) = \sup_{\mathbf{q}\in \mathring{\mathbf{L}}_2(\Omega)} \mathsf{I}(\mathbf{q}) \tag{251}$$

for the **pressure** function. Note that the functional I has no explicit integral-type form and is defined as a supremum–functional. The solvability of this problem follows from the coercivity condition (250). To prove (250) we apply Lemma 2.

Coercivity of the variational problem for the pressure function

Indeed, by Lemma 2 for any $\mathbf{q} \in \mathring{\mathbf{L}}_2(\Omega)$ we find $\mathbf{v_q} \in \mathbf{W_0}$ such that

$$\mathbf{div}\mathbf{v_q} = \mathbf{q} \quad \text{and} \quad \|\nabla \mathbf{v_q}\| \le \kappa_2 \|\mathbf{q}\|.$$

Take $\mathbf{v} = \mu \mathbf{v_q} + \mathbf{u_0}$ and recall that $\mathbf{div}\mathbf{u_0} = \mathbf{0}$. Then,

$$\inf_{\mathbf{v} \in \mathbf{W_0}+\mathbf{u_0}} \mathbf{L}(\mathbf{v},\mathbf{q}) \le \int_\Omega \Big(\frac{\nu}{2} |\nabla(\mu \mathbf{v_q}+\mathbf{u_0})|^2 - \mathbf{f}\cdot(\mu \mathbf{v_q}+\mathbf{u_0}) - \mathbf{q}\mathbf{div}(\mu \mathbf{v_q}+\mathbf{u_0}) \Big) \mathbf{dx} \le$$

$$\le \int_\Omega \left(\frac{\nu}{2} |\nabla \mathbf{u_0}|^2 - \mathbf{f}\cdot\mathbf{u_0} \right) \mathbf{dx} + \mu(\nu \|\nabla \mathbf{u_0}\| + \mathbf{C_\Omega}\|\mathbf{f}\|)\|\nabla \mathbf{v_q}\| +$$

$$+ \frac{\nu\mu^2}{2} \|\nabla \mathbf{v_q}\|^2 - \mu \|\mathbf{q}\|^2 \le \int_\Omega \left(\frac{\nu}{2} |\nabla \mathbf{u_0}|^2 - \mathbf{f}\cdot\mathbf{u_0} \right) \mathbf{dx} +$$

$$+ \mu(\nu \|\nabla \mathbf{u_0}\| + \mathbf{C_\Omega}\|\mathbf{f}\|)\kappa_2 \|\mathbf{q}\| + \mu \left(\frac{\nu\mu\kappa_2^2}{2} - 1 \right) \|\mathbf{q}\|^2,$$

where $\mathbf{C_\Omega}$ is a constant in the Friederichs inequality.

We see that

$$\begin{aligned} \mathsf{I}(\mathbf{q}) \leq \quad & c_1(\mathbf{u_0}, \mathbf{f}, \boldsymbol{\nu}) + \mu(\boldsymbol{\nu}\|\nabla \mathbf{u_0}\| + \mathbf{C_\Omega}\|\mathbf{f}\|)\boldsymbol{\kappa_2}\|\mathbf{q}\| + \\ & + \mu\left(\frac{\boldsymbol{\nu}\mu\boldsymbol{\kappa}_2^2}{\mathbf{2}} - \mathbf{1}\right)\|\mathbf{q}\|^2. \end{aligned}$$

Set here $\mu = \frac{1}{\boldsymbol{\nu}\boldsymbol{\kappa}_2^2}$. Then

$$\inf_{\mathbf{v}\in\mathbf{W_0}+\mathbf{u_0}} \mathsf{L}(\mathbf{v}, \mathbf{q}) \quad \leq c_1 + c_2\|\mathbf{q}\| - \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\nu}\boldsymbol{\kappa}_2^2}\|\mathbf{q}\|^2 \to -\infty \quad \text{as } \|\mathbf{q}\| \to +\infty.$$

Thus, we observe that the constant $\boldsymbol{\kappa}_2$ arises in the quadratic term that provides the required coercivity property of the pressure functional.

Estimates of the distance to the set of solenoidal fields

Now we are concerned with the estimates of the **distance between a function $\widehat{\mathbf{v}} \in \mathbf{H}^1$ and the space of solenoidal functions**.

Estimates in $\mathbf{L}_2$-norm. An estimate of the distance between $\widehat{\mathbf{v}}$ and the space

$$\mathbf{J}_2^1(\Omega) := \left\{ \mathbf{v} \in \mathbf{W}_2^1(\Omega) \mid \mathbf{div v} = 0 \right\}$$

in $\mathbf{L}_2$–norm follow from the solvability of the Dirichlét problem for the Lapalce operator. It is as follows:

$$\inf_{\mathbf{v}_0 \in \mathbf{J}_2^1} \|\widehat{\mathbf{v}} - \mathbf{v}_0\| \leq \mathbf{C_F} \|\mathbf{div}\widehat{\mathbf{v}}\|,$$

where $\mathbf{C_F}$ is the constant in the Friederichs inequality.

Proof. Indeed, since the problem

$$\mathbf{\Delta}\phi = \mathbf{f},$$

has a solution $\phi \in \mathring{\mathbf{W}}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega})$ for any $\mathbf{f} \in \mathbf{L_2(\Omega)}$, we conclude that for any $\mathbf{f}$ there exists $\mathbf{v_f} = \nabla\phi$ such that

$$\mathbf{div v_f} = \mathbf{f} \qquad \text{and} \qquad \|\mathbf{v_f}\| \le \mathbf{C_F}\|\mathbf{f}\|.$$

Set $\mathbf{f} = \mathbf{div}\widehat{\mathbf{v}}$. Then,

$$\mathbf{div}(\mathbf{v_f} - \widehat{\mathbf{v}}) = \mathbf{0},$$

so that $\mathbf{v_0} = \mathbf{v_f} - \widehat{\mathbf{v}}$ belongs to $\mathbf{J}^{\mathbf{1}}_{\mathbf{2}}$ and we observe that

$$\|\widehat{\mathbf{v}} - \mathbf{v_0}\| \le \mathbf{C_F}\|\mathbf{div}\widehat{\mathbf{v}}\|.$$

Estimate in $\mathbf{H^1}$-norm. Let now $\widehat{\mathbf{v}} \in \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}$. Set $\mathbf{f} = \mathbf{div}\widehat{\mathbf{v}}$. Since

$$\int_{\Omega} \mathbf{div}\widehat{\mathbf{v}}\, \mathbf{dx} = \int_{\partial\Omega} \mathbf{v} \cdot \boldsymbol{\nu}\, \mathbf{ds} = \mathbf{0},$$

we see that $\mathbf{f} \in \overset{\circ}{\mathbf{L}}{}_{\mathbf{2}}(\boldsymbol{\Omega})$. Then, by Lemma 2, one can find $\mathbf{u_f} \in \mathbf{W_0}$ such that

$$\mathbf{div}\mathbf{u_f} = \mathbf{div}\widehat{\mathbf{v}}, \quad \text{and} \quad \|\nabla \mathbf{u_f}\| \le \boldsymbol{\kappa}_{\mathbf{2}}(\boldsymbol{\Omega}) \|\mathbf{div}\widehat{\mathbf{v}}\|.$$

In other words, there exists a solenoidal field $\mathbf{w_0} = (\widehat{\mathbf{v}} - \mathbf{u_f}) \in \mathbf{W_0}$ such that

$$\|\nabla(\widehat{\mathbf{v}} - \mathbf{w_0})\| = \|\nabla \widehat{\mathbf{u}}_{\mathbf{f}}\| \le \boldsymbol{\kappa}_{\mathbf{2}}(\boldsymbol{\Omega}) \|\mathbf{div}\widehat{\mathbf{v}}\|.$$

This fact can be presented in another form

$$\inf_{\mathbf{v}\in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega)} \|\nabla(\widehat{\mathbf{v}} - \mathbf{v})\| \leq \kappa_2(\Omega)\|\mathbf{div}\widehat{\mathbf{v}}\|. \tag{252}$$

Thus, for the functions with zero traces the distance to $\overset{\circ}{\mathbf{J}}{}^1_2(\Omega)$ in a strong norm is also measured via $\|\mathbf{div}\widehat{\mathbf{v}}\|$, but with a different factor: $\kappa_2(\Omega)$.

Comments on the value of C

Note that $\mathbf{C}$ can be estimated throughout the constant $\mathbf{C_F}$ and the constant $\mathbf{C_P}$ in the Poincare inequality. Indeed,

$$\mathbf{C} = \inf_{\mathbf{q} \in \mathring{\mathbf{L}}_2(\Omega),\, \mathbf{q} \neq 0} \mathcal{E}(\mathbf{q}),$$

$$\mathcal{E}(\mathbf{q}) = \sup_{\mathbf{w} \in \mathbf{W_0},\, \mathbf{w} \neq 0} \frac{\int_\Omega \mathbf{q}\, \mathbf{div w}\, \mathbf{dx}}{\|\mathbf{q}\|\, \|\nabla \mathbf{w}\|}.$$

For $\mathbf{q} \in \widetilde{\mathbf{W}}\,(\Omega) := \mathring{\mathbf{L}}_2(\Omega) \cap \mathbf{W}_2^1(\Omega)$ we have

$$\mathcal{E}(\mathbf{q}) = \sup_{\mathbf{w} \in \mathbf{W_0},\, \mathbf{w} \neq 0} \frac{\int_\Omega \nabla \mathbf{q} \cdot \mathbf{w}\, \mathbf{dx}}{\|\mathbf{q}\|\, \|\nabla \mathbf{w}\|} \leq \frac{\|\nabla \mathbf{q}\|}{\|\mathbf{q}\|} \sup_{\mathbf{w} \in \mathbf{W_0},\, \mathbf{w} \neq 0} \frac{\|\mathbf{w}\|}{\|\nabla \mathbf{w}\|}$$

$$\leq \mathbf{C_F} \frac{\|\nabla \mathbf{q}\|}{\|\mathbf{q}\|}.$$

Let $\mathbf{C_P}$ be the smallest constant in the inequality

$$\|\mathbf{q}\| \leq \mathbf{C_P}\|\nabla \mathbf{q}\|, \quad \mathbf{q} \in \widetilde{\mathbf{W}}(\Omega),$$

i.e.,

$$\inf_{\mathbf{q}\in\widetilde{\mathbf{W}}(\Omega),\, \mathbf{q}\neq\mathbf{0}} \frac{\|\nabla \mathbf{q}\|}{\|\mathbf{q}\|} = \frac{1}{\mathbf{C_P}}.$$

Then

$$\mathbf{C} = \inf_{\mathbf{q}\in\mathring{\mathbf{L}}_2(\Omega),\, \mathbf{q}\neq\mathbf{0}} \mathcal{E}(\mathbf{q}) \leq \inf_{\mathbf{q}\in\widetilde{\mathbf{W}}(\Omega),\, \mathbf{q}\neq\mathbf{0}} \mathcal{E}(\mathbf{q}) \leq \frac{\mathbf{C_F}}{\mathbf{C_P}}.$$

LBB-condition can be written in the form

$$\|\mathbf{p}\| \leq \mathbf{C}^{-1} \, \mathbf{[} \nabla \mathbf{p} \mathbf{]} \quad \forall \ \mathbf{p} \in \mathring{\mathbf{L}}_2(\Omega),$$

what amounts

$$\mathbf{C} \leq \frac{\mathbf{[} \nabla \mathbf{p} \mathbf{]}}{\|\mathbf{p}\|}$$

we see the meaning of this constant: **C is the infimum of $\mathbf{H}^{-1}$ norms of functions such that $\|\mathbf{p}\| = 1$ and $\int_\Omega \mathbf{p} \, \mathbf{dx} = \mathbf{0}$.**

Proposition 1

If $\Omega \in \mathbb{R}^n$ then

$$\frac{\|\nabla \mathbf{p}\|_{(-1)}}{\|\mathbf{p}\|} \leq \mathbf{n} \qquad \forall \mathbf{p} \in \mathbf{L}_2(\Omega).$$

Proof.

$$\sup_{\substack{w \in W_0 \\ w \neq 0}} \frac{\int_\Omega p \, \mathrm{div} w \, dx}{\|\nabla w\|} =$$

$$\sup_{\substack{w \in W_0 \\ w \neq 0}} \frac{\sum_{t=1}^{n} \int_\Omega p \, w_{t,t} \, dx}{\|\nabla w\|} \leq \sum_{t=1}^{n} \sup_{\substack{w_t \in \overset{\circ}{H}{}^1 \\ w_t \neq 0}} \frac{\int_\Omega p \, w_{t,t} \, dx}{\|\nabla w\|} .$$

Since

$$\|\nabla \mathbf{w}\|^2 = \int_\Omega \Big(\sum_{t,s=1,n} \mathbf{w}^2_{t,s} \Big)\, \mathbf{dx} \geq \int_\Omega \mathbf{w}^2_{t,t}\, \mathbf{dx} \quad \forall \mathbf{t} = 1, 2, \ldots \mathbf{n}$$

we have

$$\sup_{\substack{\mathbf{w} \in \mathbf{W_0} \\ \mathbf{w} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\, \mathbf{div w}\, \mathbf{dx}}{\|\nabla \mathbf{w}\|} \leq \sum_{t=1}^{n} \sup_{\substack{\mathbf{w_t} \in \mathring{\mathbf{H}}^1 \\ \mathbf{w_t} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\, \mathbf{w}_{t,t}\, \mathbf{dx}}{\|\mathbf{w}_{t,t}\|} \leq$$

$$\leq \sum_{t=1}^{n} \sup_{\substack{\eta \in \mathbf{L_2} \\ \eta \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\, \eta\, \mathbf{dx}}{\|\eta\|} = \sum_{t=1}^{n} \|\mathbf{p}\| = \mathbf{n}\, \|\mathbf{p}\| \ .$$

Proposition 2

If $\mathbf{n = 1}$ then $\mathbf{C} = 1$.

Let $\mathbf{\Omega = (a, b)}$. Due to Proposition 1 we see that $\mathbf{C} \leq 1$. Let $\mathbf{p}$ be an arbitrary function from the set $\overset{\circ}{L}_2(\Omega)$. Then, the function

$$\mathbf{w^{(p)}} = \int_{\mathbf{a}}^{\mathbf{x}} \mathbf{p}\,\mathbf{dx} \in \mathbf{W_0}\,.$$

Really, $\mathbf{w^{(p)}(a) = 0}$, $\mathbf{w^{(p)}(b) = \int_b^a p dx = 0}$ and $\mathbf{w^{(p)\prime} = p} \in \mathbf{L_2(a, b)}$. Thus,

$$\sup_{\substack{\mathbf{w} \in \mathbf{W_0} \\ \mathbf{w} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\,\mathbf{w}'\,\mathbf{dx}}{\|\mathbf{w}'\|} \geq \frac{\int_\Omega \mathbf{p}\,\mathbf{w}^{(p)\prime}\,\mathbf{dx}}{\left\|\mathbf{w}^{(p)\prime}\right\|} = \frac{\int_\Omega \mathbf{p}^2\,\mathbf{dx}}{\|\mathbf{p}\|} = \|\mathbf{p}\|$$

Thus, $\mathbf{C} \geq 1$ and we arrive at the required result.

These estimates give a certain presentation on the value of $\mathbf{C}$. However, we are mainly interested in the estimate from below, what imposes a task more complicated than the finding the constant in the Friederichs inequality.
In principle, one could determine $\mathbf{C}$ by the following arguments. Let $\mathbf{w_p} \in \mathbf{W_0}$ be a function such that

$$\Delta \mathbf{w_p} = \nabla \mathbf{p}, \quad \mathbf{w_p} = \mathbf{0} \ \text{ on } \ \partial\Omega \, .$$

Then,

$$-\int_\Omega \nabla \mathbf{w_p} : \nabla \mathbf{v} \, \mathbf{dx} = \int_\Omega \nabla \mathbf{p} \cdot \mathbf{v} \, \mathbf{dx} \quad \forall \mathbf{v} \in \mathbf{W_0}$$

and, thus, we have

$$\int_\Omega | \nabla \mathbf{w_p} |^2 \, \mathbf{dx} = \int_\Omega \mathbf{p} \, \mathbf{div} \mathbf{w_p} \, \mathbf{dx} \quad \forall \mathbf{v} \in \mathbf{W_0} \, .$$

Therefore,

$$\mathbf{C} := \inf_{\substack{\mathbf{p} \in \mathring{\mathbf{L}}_2(\Omega) \\ \mathbf{p} \neq \mathbf{0}}} \sup_{\substack{\mathbf{w} \in \mathbf{W}_0 \\ \mathbf{w} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\,\mathrm{div}\mathbf{w}\,dx}{\|\mathbf{p}\|\,\|\nabla \mathbf{w}\|} \geq$$

$$\geq \inf_{\substack{\mathbf{p} \in \mathring{\mathbf{L}}_2(\Omega) \\ \mathbf{p} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\,\mathrm{div}\mathbf{w}_\mathbf{p} dx}{\|\mathbf{p}\|\,\|\nabla \mathbf{w}_\mathbf{p}\|} = \inf_{\substack{\mathbf{p} \in \mathring{\mathbf{L}}_2(\Omega) \\ \mathbf{p} \neq \mathbf{0}}} \frac{\|\nabla \mathbf{w}_\mathbf{p}\|}{\|\mathbf{p}\|}\,.$$

Thus, finding $\mathbf{C}$ requires the minimization of this quotient with respect to all $\mathbf{p} \in \mathring{\mathbf{L}}_2(\Omega)$, where $\mathbf{w}_\mathbf{p}$ is taken as the solution of the above defined linear problem. Certainly, such a task (for some Ω) might be solved by only analytical methods.

C for square domains

We will use the above relation to minimize the quotient on a subspace of $\overset{\circ}{L}_2(\Omega)$ what may give a presentation on the value of $\mathbf{C}$.
Let

$$\mathbf{\Omega} = \mathbb{Q} := \{\mathbf{x} \in \mathbb{R}^{\mathbf{n}} \mid\mid \mathbf{x_i} \in (-\pi, \pi),\ \mathbf{i} = \mathbf{1}, \mathbf{2}, ...\mathbf{n}\}.$$

We are interested in the value of the quotient

$$\inf_{\mathbf{p} \in \overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\mathbf{\Omega})} \frac{\mathbf{[}\nabla \mathbf{p}\mathbf{]}}{\|\mathbf{p}\|_{\mathbb{Q}}}.$$

Represent $\mathbf{p}$ as a series with respect to the trial functions

$$\mathbf{p}_{\mathbf{ij}}^{(1)} = \sin \mathbf{ix} \sin \mathbf{jy}, \quad \mathbf{p}_{\mathbf{ij}}^{(2)} = \sin \mathbf{ix} \cos \mathbf{jy},$$
$$\mathbf{p}_{\mathbf{ij}}^{(3)} = \cos \mathbf{ix} \sin \mathbf{jy}, \quad \mathbf{p}_{\mathbf{ij}}^{(4)} = \cos \mathbf{ix} \cos \mathbf{jy},$$

where $\mathbf{i}, \mathbf{j} = \mathbf{0}, \mathbf{1}, \mathbf{2}, ...$

Then

$$\mathbf{p(x,y)} = \sum_{\mathbf{i,j=0}}^{\infty} \sum_{\mathbf{s=1}}^{\mathbf{4}} \mathbf{a_{ij}^{(s)} p_{ij}^{(s)}} \, .$$

Here, the first nonzero coefficients are

$$\mathbf{a_{00}^{(4)}} = \frac{\mathbf{1}}{\mathbf{4\pi^2}} \int_{\Omega} \mathbf{p \, dxdy} \, , \quad \mathbf{a_{i0}^{(2)}} = \frac{\mathbf{1}}{\mathbf{2\pi^2}} \int_{\Omega} \mathbf{p} \, \sin \mathbf{ix \, dxdy} \, ,$$

$$\mathbf{a_{0j}^{(3)}} = \frac{\mathbf{1}}{\mathbf{2\pi^2}} \int_{\Omega} \mathbf{p} \, \sin \mathbf{jy \, dxdy} \, ,$$

$$\mathbf{a_{i0}^{(4)}} = \frac{\mathbf{1}}{\mathbf{2\pi^2}} \int_{\Omega} \mathbf{p} \, \cos \mathbf{ix \, dxdy} \, ,$$

$$\mathbf{a_{0j}^{(2)}} = \frac{\mathbf{1}}{\mathbf{2\pi^2}} \int_{\Omega} \mathbf{p} \, \cos \mathbf{jy \, dxdy} \, ,$$

Other coefficients are as follows:

$$\mathbf{a_{ij}^{(1)}} = \frac{\mathbf{1}}{\pi^2} \int_{\Omega} \mathbf{p} \sin \mathbf{ix} \sin \mathbf{jy} \, \mathbf{dxdy} \,,$$

$$\mathbf{a_{ij}^{(2)}} = \frac{\mathbf{1}}{\pi^2} \int_{\Omega} \mathbf{p} \sin \mathbf{ix} \cos \mathbf{jy} \, \mathbf{dxdy} \,,$$

$$\mathbf{a_{ij}^{(3)}} = \frac{\mathbf{1}}{\pi^2} \int_{\Omega} \mathbf{p} \cos \mathbf{ix} \sin \mathbf{jy} \, \mathbf{dxdy} \,,$$

$$\mathbf{a_{ij}^{(4)}} = \frac{\mathbf{1}}{\pi^2} \int_{\Omega} \mathbf{p} \cos \mathbf{ix} \cos \mathbf{jy} \, \mathbf{dxdy} \,.$$

We have

$$\|\mathbf{p}\|_{\mathbb{Q}}^{\mathbf{2}} = \pi^2 \sum_{\mathbf{i,j=0}}^{\infty} \lambda_{\mathbf{ij}} \left[\left(\mathbf{a_{ij}^{(1)}} \right)^2 + \left(\mathbf{a_{ij}^{(2)}} \right)^2 + \left(\mathbf{a_{ij}^{(3)}} \right)^2 + \left(\mathbf{a_{ij}^{(4)}} \right)^2 \right] ,$$

where $\lambda_{00} = 0$, $\lambda_{01} = 2$, $\lambda_{10} = 2$ and $\lambda_{ij} = 1$ for all $\mathbf{i, j} \geq \mathbf{1}$.

Let us take a finite number of elements in the Fourier series for $\mathbf{p}$:

$$\mathbf{p} = \sum_{\mathbf{i,j=0}}^{\mathbf{N}} \sum_{\mathbf{s=1}}^{\mathbf{4}} \mathbf{a}_{\mathbf{ij}}^{(\mathbf{s})} \mathbf{p}_{\mathbf{ij}}^{(\mathbf{s})} ,$$

where $\mathbf{a}_{\mathbf{ij}}^{(\mathbf{s})}$ are the above defined coefficients. Since

$$\| \nabla \mathbf{p} \| = \sup_{\mathbf{v} \in \mathbf{W_0}} \frac{\int_{\Omega} \mathbf{p} \, \mathrm{div} \mathbf{v} \, \mathbf{dx}}{\| \nabla \mathbf{v} \|_{\mathbb{Q}}}$$

we need to introduce the system of trial functions in $\mathbb{W}_0(\mathbb{Q})$. It is given by the system of eigenfunctions for the problem

$$\boldsymbol{\Delta} \, \mathbf{w} = \mu \, \mathbf{w} \qquad \mathbf{w}|_{\partial \mathbb{Q}} = \mathbf{0} .$$

This system is

$$\phi_{\alpha\beta} = \sin \frac{\alpha}{2}(x+\pi) \ \sin \frac{\beta}{2}(y+\pi)\,.$$

In this case,

$$\phi_{\alpha\beta,1} = \frac{\alpha}{2} \cos \frac{\alpha}{2}(x+\pi) \ \sin \frac{\beta}{2}(y+\pi)\,,$$
$$\phi_{\alpha\beta,2} = \frac{\beta}{2} \sin \frac{i}{2}(x+\pi) \ \cos \frac{\beta}{2}(y+\pi)\,.$$

Take a finite number $\mathbf{M}$ of basic functions in the representation of $\mathbf{v}$, namely we set

$$\mathbf{v} = \mathbf{v}^{\mathbf{M}} = (\mathbf{v}_{\mathbf{1}}^{\mathbf{M}}, \mathbf{v}_{\mathbf{2}}^{\mathbf{M}}),\ \mathbf{v}_{\mathbf{1}}^{\mathbf{M}} = \sum_{\alpha,\beta=\mathbf{1}}^{\mathbf{M}} \mathbf{b}_{\alpha\beta}\phi_{\alpha\beta},\ \mathbf{v}_{\mathbf{2}}^{\mathbf{M}} = \sum_{\alpha,\beta=\mathbf{1}}^{\mathbf{M}} \mathbf{c}_{\alpha\beta}\phi_{\alpha\beta}\,.$$

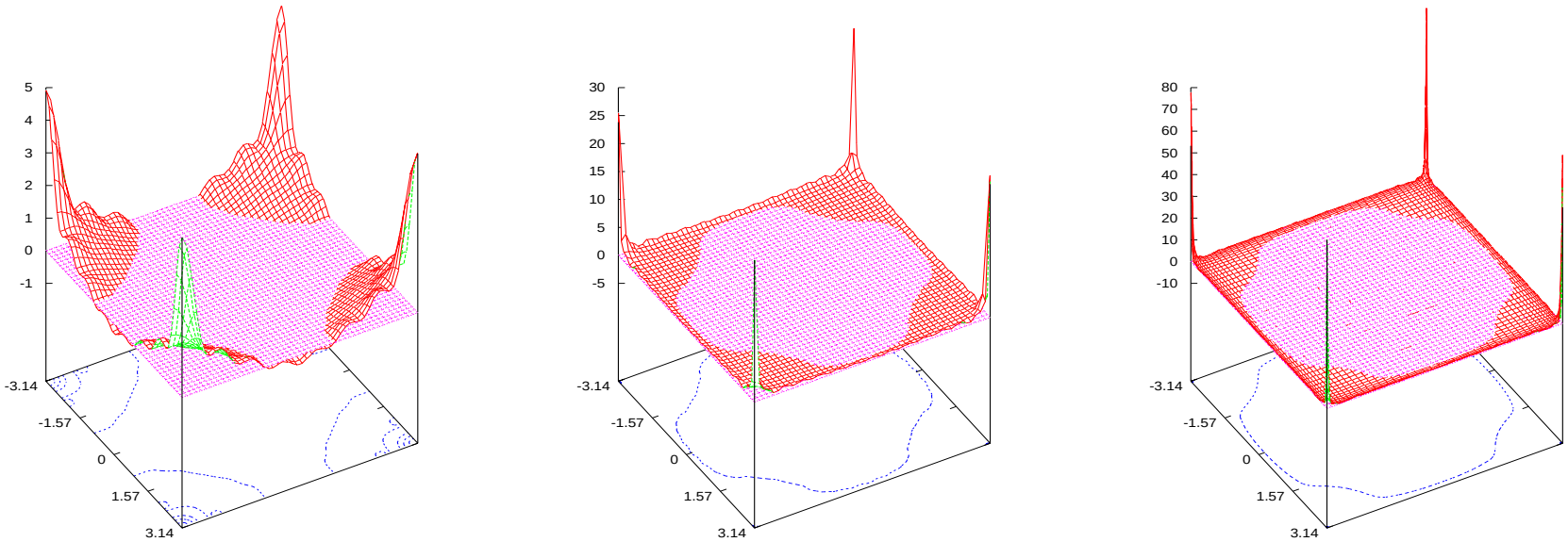
The set of all such functions we denote $\mathbf{W}_{\mathbf{0}}^{\mathbf{M}}$. In this case, we can obtain a lower bound for the required norm. Really, we have

$$\mathbf{]}\nabla \mathbf{p}\mathbf{[}^{(\mathbf{M})} := \sup_{\mathbf{v}^{\mathbf{M}} \in \mathbf{W}_0^{\mathbf{M}}} \frac{\int_{\Omega} \mathbf{p}\, \mathbf{div} \mathbf{v}^{\mathbf{M}}\, \mathbf{dx}}{\|\nabla \mathbf{v}^{\mathbf{M}}\| \mathbb{Q}} \leq \mathbf{]}\nabla \mathbf{p}\mathbf{[} = \sup_{\mathbf{v} \in \mathbf{W_0}} \frac{\int_{\Omega} \mathbf{p}\, \mathbf{div} \mathbf{v}\, \mathbf{dx}}{\|\nabla \mathbf{v}\|_{\mathbb{Q}}} .$$

Thus, we may hope to estimate the value of the quotient

$$\inf_{\mathbf{p} \in \overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\Omega)} \frac{\mathbf{]}\nabla \mathbf{p}\mathbf{[}}{\|\mathbf{p}\|_{\mathbb{Q}}} .$$

by taking $\mathbf{N}, \mathbf{M} \to +\infty$, $\mathbf{M} = \kappa \mathbf{N}$ κ is essentially larger than 1 (typically 8-20). Numerical results for different $\mathbf{N}$ are exposed below.



Minimizer p_n for n =8, 36 and 120.

Deviation estimates for the Stokes problem

In order to clarify the main ideas of our approach we rewrite the classical Stokes system in a somewhat different form:

$$\mathbf{div}\boldsymbol{\sigma} = \nabla \mathbf{p} - \mathbf{f} \quad \text{in } \Omega, \tag{253}$$

$$\mathbf{div}\mathbf{u} = \mathbf{0} \quad \text{in } \Omega, \tag{254}$$

$$\boldsymbol{\sigma} = \nu \nabla \mathbf{u} \quad \text{in } \Omega, \tag{255}$$

$$\mathbf{u} = \mathbf{0} \quad \text{on } \partial\Omega. \tag{256}$$

This system involves one additional variable $\boldsymbol{\sigma}$ that corresponds to the field of stresses. Now we may regard the Stokes problem as the problem of finding a triplet of functions $(\mathbf{u}, \boldsymbol{\sigma}, \mathbf{p})$.

Primal and Dual Problems

Functional formulations of the above problem are given in natural "energy" set for this velocity–stress-pressure setting, which is

$$\mathcal{E} := \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) \times \mathbf{\Sigma} \times \overset{\circ}{\mathbf{L}}{}_{2}(\Omega).$$

Problem $\mathcal{P}$**.** Find $\mathbf{u} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega)$ such that

$$\mathbf{J}(\mathbf{u}) \leq \mathbf{J}(\mathbf{v}) \quad \text{for all} \quad \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega)\,,$$

where

$$\mathbf{J}(\mathbf{v}) = \int_{\Omega} \left(\frac{\nu}{2} |\nabla \mathbf{v}|^2 - \mathbf{f} \cdot \mathbf{v} \right) \mathbf{dx}.$$

We denote the exact lower bound of this problem by $inf\,\mathcal{P}$.

Introduce the Lagrangian $\mathbf{L} : \overset{\circ}{\mathbf{J}}{}^1_2(\Omega) \times \boldsymbol{\Sigma}(\Omega) \to \mathbb{R}$:

$$\mathbf{L}(\mathbf{v}, \tau) = \int_\Omega \left(\tau : \nabla \mathbf{v} - \frac{1}{2\nu} |\tau|^2 \right) \mathbf{dx} - \int_\Omega \mathbf{f} \mathbf{v} \, \mathbf{dx}$$

that together with Problem $\mathcal{P}$ generates the dual problem

$$\sup_{\tau \in \boldsymbol{\Sigma}} \inf_{\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega)} \mathbf{L}(\mathbf{v}, \tau)$$

which is **Problem $\mathcal{P}^*$:** find $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_\mathbf{f}$ such that

$$\mathbf{I}^*(\boldsymbol{\sigma}) = \sup_{\tau \in \boldsymbol{\Sigma}_\mathbf{f}} \mathbf{I}^*(\tau), \quad \mathbf{I}^*(\tau) = -\frac{1}{2\nu} \int_\Omega |\tau|^2 \, \mathbf{dx}$$

where

$$\boldsymbol{\Sigma}_\mathbf{f} := \left\{ \tau \in \boldsymbol{\Sigma}(\Omega) \; || \int_\Omega \tau : \nabla \mathbf{w} \, \mathbf{dx} = \int_\Omega \mathbf{f} \mathbf{w} \mathbf{dx} \;\; \text{for all } \mathbf{w} \in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega) \right\}.$$

From the general theorems of convex analysis it follows

Theorem (1)

There exists a unique minimizer $\mathbf{u}$ *of problem* $\mathcal{P}$ *and unique maximizer* $\boldsymbol{\sigma}$ *of problem* $\mathcal{P}^*$*. These two functions meet the equalities*

$$\mathbf{I}^*(\boldsymbol{\sigma}) = \sup \mathcal{P}^* = \inf \mathcal{P} = \mathbf{I}(\mathbf{u}), \tag{257}$$

$$\boldsymbol{\sigma} = \boldsymbol{\nu} \nabla \mathbf{u}\,. \tag{258}$$

Basic error estimate

The basic error relation for the Stokes problem is given by the following theorem (S. Repin, 2002).

Theorem (2)

For any $\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega)$ *and any* $\tau_f \in \boldsymbol{\Sigma}_{\mathbf{f}}$, *we have*

$$\nu \left\| \nabla(\mathbf{v}-\mathbf{u}) \right\|^2 + \frac{1}{\nu} \left\| \tau_{\mathbf{f}} - \sigma \right\|^2 = 2 \left(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\tau_{\mathbf{f}}) \right). \quad (259)$$

Proof of Theorem 2

Since $\mathbf{u}$ is the solution of problem $\mathcal{P}$, we obtain

$$\mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) = \int_\Omega \left(\frac{\nu}{2} \mid \nabla \mathbf{v} \mid^2 - \frac{\nu}{2} \mid \nabla \mathbf{u} \mid^2 - \mathbf{f} \cdot (\mathbf{v} - \mathbf{u}) \right) \mathbf{dx} =$$

$$= \int_\Omega \left(\frac{\nu}{2} \mid \nabla (\mathbf{v} - \mathbf{u}) \mid^2 + \nu \nabla \mathbf{u} : \nabla (\mathbf{v} - \mathbf{u}) - \mathbf{f} \cdot (\mathbf{v} - \mathbf{u}) \right) \mathbf{dx} =$$

$$= \frac{\nu}{2} \int_\Omega \mid \nabla (\mathbf{v} - \mathbf{u}) \mid^2 \mathbf{dx} \qquad \text{for all } \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega).$$

Since $\mathbf{J}(\mathbf{u}) = \inf \mathcal{P}$, we conclude that

$$\frac{\nu}{2} \| \nabla (\mathbf{v} - \mathbf{u}) \|^2 = \mathbf{J}(\mathbf{v}) - \inf \mathcal{P} \qquad \text{for all } \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega) \, .$$

The next step is to derive a similar relation for the dual problem. For this purpose, we note that the maximizer $\boldsymbol{\sigma}$ of problem $\mathcal{P}^*$ satisfies the relation

$$\int_{\Omega} \boldsymbol{\sigma} : (\boldsymbol{\tau}_f - \boldsymbol{\sigma})\, \mathbf{dx} = \mathbf{0} \qquad \text{for all } \boldsymbol{\tau}_f \in \boldsymbol{\Sigma}_f\,.$$

By virtue of this relation, we find that

$$\sup \mathcal{P}^* - \mathbf{I}^*(\boldsymbol{\tau}_f) = \mathbf{I}^*(\boldsymbol{\sigma}) - \mathbf{I}^*(\boldsymbol{\tau}_f) = \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\nu}} \|\boldsymbol{\tau}_f - \boldsymbol{\sigma}_f\|^2 \quad \boldsymbol{\tau}_f \in \boldsymbol{\Sigma}_f.$$

Since $\inf \mathcal{P} = \sup \mathcal{P}^*$ we sum the two equalities and obtain

$$\boldsymbol{\nu} \|\nabla(\mathbf{v} - \mathbf{u})\|^2 + \frac{\mathbf{1}}{\boldsymbol{\nu}} \|\boldsymbol{\tau}_f - \boldsymbol{\sigma}\|^2 = \mathbf{2}\,(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\boldsymbol{\tau}_f))\,.$$

Stokes problem is a particular case of the abstract problem we investigated in Lecture 5:

Find $\mathbf{u} \in \mathbf{V_0} + \mathbf{u_0}$ such that

$$(\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda w}) + \langle \ell, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

In this case $\mathbf{V_0} = \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega})$, $\mathbf{V}$ is a subspace of $\mathbf{H^1}$ containing solenoidal fields, $\mathbf{\Lambda} = \nabla$ (tensor–gradient), $\mathbf{U} = \mathbf{\Sigma}$, $\mathcal{A}\mathbf{y} = \boldsymbol{\nu}\mathbf{y}$, and

$$\langle \ell, \mathbf{w} \rangle = -\int_{\mathbf{\Omega}} \mathbf{f} \mathbf{w} \, \mathbf{dx}$$

Thus, we can apply the estimate

$$\frac{1}{2} \| \Lambda(v-u) \|^2 \leq (1+\beta) D(\Lambda v, y) + \frac{1+\beta}{2\beta} [\ell + \Lambda^* y]^2, \qquad (260)$$

where $\| y \|^2 = \int_\Omega \nu |y|^2 dx$ and

$$[\ell + \Lambda^* y] = \sup_{w \in V_0} \frac{\langle \ell + \Lambda^* y, w \rangle}{\| \Lambda w \|} = \sup_{w \in \overset{\circ}{J}{}^1_2(\Omega)} \frac{\int_\Omega (\nabla w : y - f w) dx}{\| \nabla w \|} =$$

$$\sup_{w \in \overset{\circ}{J}{}^1_2(\Omega)} \frac{\int_\Omega (\nabla w : y - f w - q \, div w) dx}{\| \nabla w \|} \leq$$

$$\leq \sup_{w \in \overset{\circ}{H}{}^1(\Omega)} \frac{\int_\Omega (\nabla w : y - f w - q \, div w) dx}{\| \nabla w \|} \qquad \forall q \in L^2(\Omega).$$

If

$$\mathbf{y} \in \mathbf{\Sigma}_{\mathrm{div}}(\Omega) := \{\mathbf{y} \in \mathbf{\Sigma} \mid \mathbf{div}\mathbf{y} \in \mathbf{L}^2(\Omega, \mathbb{R}^n)\}$$

and $\mathbf{q} \in \mathbf{H}^1$, we have

$$\sup_{\mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^1(\Omega)} \frac{\int_\Omega (\nabla \mathbf{w} : \mathbf{y} - \mathbf{f}\mathbf{w} - \mathbf{q}\mathbf{div}\mathbf{w}) d\mathbf{x}}{\| \nabla \mathbf{w} \|} = \sup_{\mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^1(\Omega)} \frac{\int_\Omega (\mathbf{f} - \nabla \mathbf{q} + \mathbf{div}\mathbf{y}) \cdot \mathbf{w} d\mathbf{x}}{\| \nabla \mathbf{w} \|}$$

Since

$$\|\mathbf{w}\| \le \mathbf{C}_\Omega \|\nabla \mathbf{w}\| = \mathbf{C}_\Omega \boldsymbol{\nu}^{-1/2} \| \nabla \mathbf{w} \|,$$

we obtain

$$\mathbf{]} \ell + \mathbf{\Lambda}^* \mathbf{y} \mathbf{[} \le \mathbf{C}_\Omega \boldsymbol{\nu}^{-1/2} \|\mathbf{f} - \nabla \mathbf{q} + \mathbf{div}\mathbf{y}\|$$

Further,

$$\mathbf{D}(\nabla\mathbf{v},\mathbf{y}) = \int_\Omega \left(\frac{1}{2}\nu\nabla\mathbf{v} : \nabla\mathbf{v} + \frac{1}{2}\nu^{-1}\mathbf{y} : \mathbf{y} - \nabla\mathbf{v} : \mathbf{y}\right)\mathbf{dx} =$$

$$= \frac{1}{2\nu}\|\mathbf{y} - \nu\nabla\mathbf{v}\|^2.$$

Now, from (260) we obtain

$$\frac{\nu}{2}\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \le (1+\beta)\frac{1}{2\nu}\|\mathbf{y}-\nu\nabla\mathbf{v}\|^2 + \frac{1+\beta}{2\beta\nu}\mathbf{C}_\Omega^2\|\mathbf{f}-\nabla\mathbf{q}+\mathbf{divy}\|^2,$$

or

$$\nu^2\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \le (1+\beta)\|\mathbf{y}-\nu\nabla\mathbf{v}\|^2 + \frac{1+\beta}{\beta}\mathbf{C}_\Omega^2\|\mathbf{f}-\nabla\mathbf{q}+\mathbf{divy}\|^2.$$

Deviation estimate for solenoidal approximations

By the minimization with respect to $\boldsymbol{\beta}$ we derive the first basic estimate for the Stokes problem:

$$\boxed{\nu\|\nabla(u-\mathbf{v})\| \leq \|\mathbf{y}-\nu\nabla\mathbf{v}\| + \mathbf{C}_\Omega\|\mathbf{f}-\nabla\mathbf{q}+\mathbf{divy}\|.} \qquad (261)$$

Here $\mathbf{v}$ is **any** conforming approximation of $\mathbf{u}$ and $\mathbf{y}$ is **any** tensor–function in $\boldsymbol{\Sigma}_{\mathrm{div}}(\Omega)$ and $\mathbf{q} \in \mathbf{H}^1$ is an "image" of the pressure function.
This and the next estimate for non-solenoidal approximations has been derived in '99, English translation is presented in S. Repin. A posteriori estimates for the Stokes problem, *J. Math. Sci. (New York)*, **109** (2002).

Non-solenoidal approximations

If the function $\widehat{\mathbf{v}} \in \mathbf{V_0} + \mathbf{u_0}$ does not satisfy the incompressibility condition, then the estimate of its deviation from $\mathbf{u}$ can be obtained as follows.
By Lemma 2 for the function $\widehat{\mathbf{v}}_0 := \widehat{\mathbf{v}} - \mathbf{u_0}$ one can find a function $\mathbf{w_0} \in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega)$ such that

$$\|\nabla(\widehat{\mathbf{v}}_0 - \widehat{\mathbf{w}}_0)\| \leq \kappa_2(\Omega)\|\mathbf{div}\widehat{\mathbf{v}}_0\|.$$

Then,

$$\begin{aligned}\nu\|\nabla(\mathbf{u} - \widehat{\mathbf{v}})\| &= \nu\|\nabla(\mathbf{u} - \widehat{\mathbf{v}}_0 - \mathbf{u_0})\| \leq \\ &\leq \nu\|\nabla(\mathbf{u} - (\widehat{\mathbf{w}}_0 + \mathbf{u_0}))\| + \nu\|\nabla(\widehat{\mathbf{v}}_0 - \widehat{\mathbf{w}}_0)\|.\end{aligned}$$

Use (261) to estimate the first norm in the right–hand side of this inequality.

We obtain

$$\nu\|\nabla(\mathbf{u}-\widehat{\mathbf{v}})\| \le \|\nu\nabla(\widehat{\mathbf{w}}_0+\mathbf{u}_0)-\mathbf{y}\| + \mathbf{C}_\Omega\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\| + \\ +\nu\|\nabla(\widehat{\mathbf{v}}_0-\widehat{\mathbf{w}}_0)\| \le \|\nu\nabla\widehat{\mathbf{v}}-\mathbf{y}\| + \\ +\mathbf{C}_\Omega\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\| + 2\nu\|\nabla(\widehat{\mathbf{v}}_0-\widehat{\mathbf{w}}_0)\|.$$

Hence, we arrive at the estimate

$$\boxed{\nu\|\nabla(\mathbf{u}-\widehat{\mathbf{v}})\| \le \|\nu\nabla(\widehat{\mathbf{v}})-\mathbf{y}\| + \mathbf{C}_\Omega\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\| + \frac{2\nu}{\mathbf{C}}\|\mathbf{div}\widehat{\mathbf{v}}\|.} \qquad (262)$$

Three terms in the right–hand side of the estimate present three natural parts of the error, namely **errors in the constitutive law, differential equation and incompressibility condition**.

Another form of the Majorant

Set $\mathbf{y} = \boldsymbol{\eta} + \mathbf{q}\mathbb{I}$, where $\mathbb{I}$ is the unit tensor and $\boldsymbol{\eta} \in \boldsymbol{\Sigma}_{\mathrm{div}}(\boldsymbol{\Omega})(\boldsymbol{\Omega})$. Then the Majorant comes in the form

$$\nu \, \|\nabla(\mathbf{u} - \widehat{\mathbf{v}})\| \leq \|\nu \nabla(\widehat{\mathbf{v}}) - \boldsymbol{\eta} - \mathbf{q}\mathbb{I}\| + \mathbf{C}_{\Omega} \|\mathbf{div}\boldsymbol{\eta} + \mathbf{f}\| + \frac{2\nu}{\mathbf{C}} \|\mathbf{div}\widehat{\mathbf{v}}\|. \qquad (263)$$

Thus, if the constants $\mathbf{c}_{\Omega}$ and $\mathbf{C}$ are known (or we know suitable upper bounds for them), then (262) and (263) provides a way of practical estimation the deviation of $\widehat{\mathbf{v}}$ from $\mathbf{u}$.

Practical implementation

To use the above estimates in practice we should select certain finite dimensional subspaces

$$\mathbf{\Sigma_k} \qquad \textbf{and} \qquad \mathbf{Q_k}$$

for the functions $\mathbf{y}$ (or $\boldsymbol{\eta}$) and $\mathbf{q}$, respectively.
Minimization of the right–hand side of the estimates with respect to $\mathbf{y}$ and $\mathbf{q}$ gives an estimate of the deviation, which will be the sharper the greater is the dimensionality of the subspaces used.

Estimates for the pressure field

Let $\mathbf{q} \in \mathring{\mathbf{L}}_2(\Omega)$ be an approximation of the pressure field $\mathbf{p}$. Then $(\mathbf{p} - \mathbf{q}) \in \mathring{\mathbf{L}}_2(\Omega)$ and the Inf-Sup condition implies the relation

$$\sup_{\mathbf{w} \in \mathbf{V_0},\, \mathbf{w} \neq \mathbf{0}} \frac{\int_\Omega (\mathbf{p} - \mathbf{q})\, \mathbf{div w}\, \mathbf{dx}}{\|\mathbf{p} - \mathbf{q}\|\ \|\nabla \mathbf{w}\|} \geq \mathbf{C}\,.$$

Thus, for any small positive ϵ there exists a nonzero function $\mathbf{w}^{\epsilon}_{\mathbf{pq}} \in \mathbf{V_0}$ such that

$$\int_\Omega (\mathbf{p} - \mathbf{q}) \mathrm{div} \mathbf{w}^{\epsilon}_{\mathbf{pq}} \mathbf{dx} \geq (\mathbf{C} - \epsilon) \|\mathbf{p} - \mathbf{q}\| \|\nabla \mathbf{w}^{\epsilon}_{\mathbf{pq}}\|.$$

Since

$$\int_\Omega \nu\,\varepsilon(\mathbf{u}) : \varepsilon(\mathbf{w}^\epsilon_{\mathbf{pq}})\,\mathbf{dx} = \int_\Omega \left(\mathbf{f}\cdot\mathbf{w}^\epsilon_{\mathbf{pq}} + \mathbf{p}\,\mathrm{div}\mathbf{w}^\epsilon_{\mathbf{pq}}\right)\,\mathbf{dx},$$

we have

$$\int_\Omega (\mathbf{p}-\mathbf{q})\mathbf{div}\,\mathbf{w}^\epsilon_{\mathbf{pq}}\,\mathbf{dx} =$$

$$= \int_\Omega \left\{\nu\varepsilon((\mathbf{u}-\widehat{\mathbf{v}}) : \varepsilon(\mathbf{w}^\epsilon_{\mathbf{pq}}) + \left(\nu\varepsilon(\widehat{\mathbf{v}}) : \varepsilon(\mathbf{w}^\epsilon_{\mathbf{pq}}) + \nabla\mathbf{q}\cdot\mathbf{w}^\epsilon_{\mathbf{pq}} - \mathbf{f}\cdot\mathbf{w}^\epsilon_{\mathbf{pq}}\right)\right\}\mathbf{dx}$$

$$= \int_\Omega \nu\varepsilon((\mathbf{u}-\widehat{\mathbf{v}}) : \varepsilon(\mathbf{w}^\epsilon_{\mathbf{pq}})\,\mathbf{dx} + \int_\Omega (\nu\varepsilon(\widehat{\mathbf{v}}-\mathbf{y}) : \varepsilon(\mathbf{w}^\epsilon_{\mathbf{pq}})\,\mathbf{dx}$$

$$+ \int_\Omega \left(\mathbf{y} : \varepsilon(\mathbf{w}^\epsilon_{\mathbf{pq}}) + \nabla\mathbf{q}\cdot\mathbf{w}^\epsilon_{\mathbf{pq}} - \mathbf{f}\cdot\mathbf{w}^\epsilon_{\mathbf{pq}}\right)\,\mathbf{dx},$$

where $\widehat{\mathbf{v}}$ is an arbitrary function in $\mathbf{W_0}+\mathbf{u_0}$ and $\mathbf{y}$ as an arbitrary tensor–valued function in $\mathbf{\Sigma}$.

Above relations lead to the estimates

$$\|\mathbf{p}-\mathbf{q}\| \le \frac{\mathbf{1}}{(\mathbf{C}-\epsilon)\|\nabla \mathbf{w}^{\epsilon}_{\mathbf{pq}}\|} \times \Big[\int_{\Omega} \big(\boldsymbol{\nu}\boldsymbol{\varepsilon}(\mathbf{u}-\widehat{\mathbf{v}}) : \boldsymbol{\varepsilon}(\mathbf{w}^{\epsilon}_{\mathbf{pq}}) + (\boldsymbol{\nu}\boldsymbol{\varepsilon}(\widehat{\mathbf{v}})-\mathbf{y}) : \boldsymbol{\varepsilon}(\mathbf{w}^{\epsilon}_{\mathbf{pq}})\big)\,\mathbf{dx} + \int_{\Omega} \big(-\mathbf{w}^{\epsilon}_{\mathbf{pq}}\cdot\mathbf{divy} + \nabla\mathbf{q}\cdot\mathbf{w}^{\epsilon}_{\mathbf{pq}} - \mathbf{f}\cdot\mathbf{w}^{\epsilon}_{\mathbf{pq}}\big)\,\mathbf{dx}\Big]$$

$$\le \frac{1}{(\mathbf{C}-\epsilon)}\big[\boldsymbol{\nu}\|\boldsymbol{\varepsilon}(\mathbf{u}-\widehat{\mathbf{v}})\| + \|\boldsymbol{\nu}\boldsymbol{\varepsilon}(\widehat{\mathbf{v}})-\mathbf{y}\| + \mathbf{C}_{\Omega}\|\mathbf{divy}+\mathbf{f}-\nabla\mathbf{q}\|\big].$$

The first term in the right–hand side of this inequality is estimated by (262).

Deviation estimate for the pressure function

Since ϵ may be taken arbitrarily small, we obtain the following estimate for the deviation from the exact pressure field:

$$\frac{1}{2}\|\mathbf{p}-\mathbf{q}\| \le \frac{\nu}{\mathbf{C}^2}\|\mathbf{div}\widehat{\mathbf{v}}\| + \tag{264}$$
$$+\frac{1}{\mathbf{C}}\|\nu\varepsilon(\widehat{\mathbf{v}})-\mathbf{y}\| + \frac{\mathbf{C}_\Omega}{\mathbf{C}}\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\|.$$

It is easy to see that the right–hand side of (264) consists of the same terms as the right–hand side of (262) and vanishes if and only if, $\widehat{\mathbf{v}}=\mathbf{u}$, $\mathbf{y}=\boldsymbol{\sigma}$ and $\mathbf{p}=\mathbf{q}$. However, in this case, the dependence of the penalty multipliers from the constant $\mathbf{C}$ is stronger.

Problems with condition div$u = \phi$.

In many cases, divergence–free condition is replaced by

$$\mathbf{divu} = \phi \quad \mathbf{in\ \Omega},$$

where ϕ is a given function in $\mathring{L}_2(\Omega)$. For such functions, we have the problem: find $\mathbf{u}$ that is equal to $\mathbf{u_0}$ on $\partial\Omega$ and

$$-\mathbf{div}\boldsymbol{\sigma} + \nabla\mathbf{p} = \mathbf{f} \quad \mathbf{in\ \Omega},$$
$$\boldsymbol{\sigma} = \nu\boldsymbol{\varepsilon}(\mathbf{u} \quad \mathbf{in\ \Omega},$$

Let $\mathbf{u}_\phi \in \mathbf{W_0}$, $\mathbf{divu}_\phi = \phi$. By setting $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}_\phi$ and $\bar{\mathbf{u}}_0 = \mathbf{u_0} - \mathbf{u}_\phi$, we present the boundary–value problem as follows: find $\bar{\mathbf{u}} \in \mathring{\mathbf{J}}_{\mathbf{2}}^{\mathbf{1}}(\mathbf{\Omega}) + \bar{\mathbf{u}}_\mathbf{0}$ such that

$$-\mathbf{div}\bar{\boldsymbol{\sigma}} + \nabla\mathbf{p} = \bar{\mathbf{f}} \quad \mathbf{in\ \Omega}, \qquad \bar{\mathbf{f}} = \mathbf{f} + \nu\mathbf{div}\boldsymbol{\varepsilon}(\mathbf{u}_\phi) \in \mathbf{H}^{-1},$$
$$\bar{\boldsymbol{\sigma}} = \nu\boldsymbol{\varepsilon}(\bar{\mathbf{u}}) \quad \mathbf{in\ \Omega}.$$

Assume that $\mathbf{u}$ is approximated by a certain $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$. Let $\mathbf{v}$ be presented in the form $\mathbf{v} = \bar{\mathbf{v}} + \mathbf{u}_\phi$. Now. we apply (262) to a "shifted" system and obtain

$$\|\varepsilon(\mathbf{u} - \mathbf{v})\| = \|\varepsilon(\bar{\mathbf{u}} - \bar{\mathbf{v}})\| \leq$$

$$\leq \|\nu\varepsilon(\bar{\mathbf{v}}) - \mathbf{y}\| + \mathbf{\rfloor}\, \mathbf{div y} + \bar{\mathbf{f}} - \nabla \mathbf{q} \,\mathbf{\lfloor} + \frac{2\nu}{\mathbf{C_{LBB}}} \|\mathbf{div}\bar{\mathbf{v}}\|.$$

Set here $\mathbf{y} = -\nu\nabla(\mathbf{u}_\phi) + \eta$, where η is a function in $\Sigma_{\mathbf{div}}$.
Then

$$\mathbf{div y} + \bar{\mathbf{f}} = -\nu \mathbf{div}\varepsilon(\mathbf{u}_\phi) + \mathbf{div}\eta + \bar{\mathbf{f}} = \mathbf{div}\eta + \mathbf{f}$$

and $\nu\varepsilon(\bar{\mathbf{v}}) - \mathbf{y} = \nu\varepsilon(\mathbf{v} - \mathbf{u}_\phi) - \mathbf{y} = \nu\varepsilon(\mathbf{v}) - \eta$. Therefore,

$$\|\varepsilon(\mathbf{u} - \mathbf{v})\| \leq$$

$$\leq \|\nu\varepsilon(\mathbf{v}) - \eta\| + \mathbf{\rfloor}\, \mathbf{div}\eta + \mathbf{f} - \nabla \mathbf{q} \,\mathbf{\lfloor} + \frac{2\nu}{\mathbf{C_{LBB}}} \|\mathbf{div v} - \phi\|.$$

Problems for almost incompressible fluids

Models of almost incompressible fluids are often used for constructing sequences of functions converging to a solution of the Stokes problem. In this case, the incompressibility condition is replaced by a penalty term: find $\mathbf{u}_\delta \in \mathbf{V}$ satisfying the integral identity

$$\int_\Omega \left(\nu \nabla \mathbf{u}_\delta : \nabla \mathbf{w} + \frac{1}{\delta} \mathbf{div} \mathbf{u}_\delta \, \mathbf{div} \mathbf{w}\right) \mathbf{dx} = \int_\Omega \mathbf{f} \cdot \mathbf{w} \, \mathbf{dx}, \quad \mathbf{w} \in \mathbf{W_0},$$

and the boundary condition $\mathbf{u}_\delta = \mathbf{u_0}$ on $\partial\Omega$.
It is not difficult to show (see, e.g.,R. Temam. NS Equations), that $\mathbf{u}_\delta$ tends to $\mathbf{u}$ (solution of the Stokes problem) in $\mathbf{H^1}$ norm and
$\mathbf{p}_\delta = -\frac{1}{\delta} \mathbf{div} \mathbf{u}_\delta \in \overset{\circ}{\mathbf{L_2}}(\mathbf{\Omega})$ converges to the respective pressure function $\mathbf{p}$ in $\mathbf{L_2}$ as $\delta \to \mathbf{0}$.

By (262) we can easily obtain an estimate of the difference between $\mathbf{u}$ and $\mathbf{u}_\delta$. Let us set in(262) $\mathbf{y} = \boldsymbol{\tau}_\delta := \boldsymbol{\nu}\nabla\mathbf{u}_\delta$ and $\mathbf{q} = \mathbf{p}_\delta = -\frac{1}{\delta}\mathbf{div}\mathbf{u}_\delta$. In this case, $\|\boldsymbol{\nu}\nabla\mathbf{u}_\delta - \boldsymbol{\tau}_\delta\| = \mathbf{0}$ and

$$\mathbf{[}\,\mathbf{div}\boldsymbol{\tau}_\delta + \mathbf{f} - \nabla\mathbf{p}_\delta\,\mathbf{]} = \\ = \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\int_\Omega (-\boldsymbol{\nu}\nabla\mathbf{u}_\delta : \nabla\mathbf{w} + \mathbf{f}\cdot\mathbf{w} + \mathbf{p}_\delta\mathbf{div}\mathbf{w})\,\mathbf{dx}}{\|\nabla\mathbf{w}\|} = \mathbf{0}.$$

Thus, we conclude that

$$\frac{1}{2}\|\nabla(\mathbf{u} - \mathbf{u}_\delta)\| \le \frac{\mathbf{1}}{\mathbf{C_{LBB}}}\|\mathbf{div}\mathbf{u}_\delta\|,$$

We observe that the deviation from the exact solution of the Stokes problem is controlled by the norm of the divergence of the regularized problem. Similar estimate can be obtained for the approximations constructed by means of the Uzawa algorithm.

Functional a posteriori estimates for the Stokes and some other problems were also derived by **nonvariational** techniques
(see S. Repin. *St.-Petersburg Math. J*, (2004)).
In particular, such estimates were derived for the **Oseen problem**

$$
\begin{aligned}
-\nu \mathbf{\Delta u} + \mathbf{div}(\mathbf{a} \otimes \mathbf{u}) = \mathbf{f} - \nabla \mathbf{p} \quad & \text{in } \Omega, \\
\mathbf{divu} = \mathbf{0} \quad & \text{in } \Omega, \\
\mathbf{u} = \mathbf{0} \quad & \text{on } \partial\Omega.
\end{aligned}
$$

Generalizations

A posteriori estimates of the above discussed type can be derived in the abstract form for the whole class of problems where a solution is seeking in a subspace.
Typically, we have the following diagram:

$$\begin{array}{ccccccc} \mathbf{H} & \xleftarrow{\mathbf{B}} & \mathbf{W_0} & \xrightarrow{\mathbf{\Lambda}} & \mathbf{U} & (\mathbf{Y}, \mathbf{Y}^*) \\ & & \Updownarrow & & & \\ \mathbf{H} & \xrightarrow{\mathbf{B}^*} & \mathbf{W_0^*} & \xleftarrow{\mathbf{\Lambda}^*} & \mathbf{U} & \end{array}$$

Basic problem. Find $\mathbf{p} \in \mathbf{H}$ and $\mathbf{u} \in \mathcal{V}_\mathbf{0}$ that satisfy the relation

$$(\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda w}) + \langle \mathbf{f} - \mathbf{B}^*\mathbf{p}, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{W_0},$$

where

$$\mathcal{V}_\mathbf{0} = \mathrm{Ker}\mathbf{B} := \{\mathbf{v} \in \mathbf{W_0} \mid \mathbf{Bv} = \mathbf{0}\}.$$

Assume that

$$\nu_1 \|\mathbf{y}\|^2 \leq (\mathcal{A}\,\mathbf{y}, \mathbf{y}) \leq \nu_2 \|\mathbf{y}\|^2, \quad \mathbf{y} \in \mathbf{U},$$

Let the operator $\mathbf{B}$ possesses the following property: there exists a constant α such that for any

$$\mathbf{g} \in \mathbf{Im}\,\mathbf{B} := \{\mathbf{z} \in \mathbf{H} \mid \exists \mathbf{v} \in \mathbf{W_0} : \mathbf{B}\mathbf{v} = \mathbf{z}\}$$

one can find $\mathbf{u_g} \in \mathbf{W_0}$ such that

$$\mathbf{B}\mathbf{u_g} = \mathbf{g} \quad \text{and} \quad \|\mathbf{u_g}\|_{\mathbf{W}} \leq \alpha \|\mathbf{g}\|.$$

Note that such a condition is a generalization of Lemma 2.

Under the above assumption we obtain an estimate of the deviation from $\mathbf{u}$.

Estimate of the deviation from u

$$\| \Lambda(u - \widehat{v}) \| \leq$$

$$\leq 2\sqrt{\nu_2}\alpha \|B\widehat{v}\| + \| \mathcal{A}\Lambda\widehat{v} - y \|_* + \frac{1}{\sqrt{\nu_1}} \| f + \Lambda^* y - B^* q \|.$$

where $\| y \| := (\mathcal{A}y, y)^{1/2}$, $\| y \|_* := (\mathcal{A}^{-1}y, y)^{1/2}$ We see that the terms of the estimate present errors in the basic relations

$$\begin{cases} \langle \Lambda^* \sigma + f - B^* p, w \rangle = 0 & \forall w \in V_0, \\ \sigma = \mathcal{A}\Lambda u, \\ Bv = 0. \end{cases}$$

For the Stokes problem $\mathbf{\Lambda v} = \nabla \mathbf{v}$, $\mathcal{A} = \boldsymbol{\nu}\mathbb{I}$, where $\mathbb{I}$ denotes the identity operator and $\mathbf{Bv} = -\mathbf{div v}$. It is easy to see that in this case $\boldsymbol{\nu}_1 = \boldsymbol{\nu}_2 = \boldsymbol{\nu}$,

$$\| \mathcal{A}\mathbf{\Lambda}\widehat{\mathbf{v}} - \mathbf{y} \|_* = \frac{\mathbf{1}}{\sqrt{\boldsymbol{\nu}}} \|\boldsymbol{\nu}\nabla \mathbf{v} - \mathbf{y}\|.$$

Since $\| \mathbf{\Lambda}(\mathbf{u} - \widehat{\mathbf{v}}) \| = \sqrt{\boldsymbol{\nu}} \|\mathbf{\Lambda}(\mathbf{u} - \widehat{\mathbf{v}})\|$, we find that the general estimate coincides with (262).

NONLINEAR MODELS IN THE THEORY OF FLUIDS

Bingham fluids

In these models

$$\mathbf{W}(\varepsilon) \;=\; \mu \left| \varepsilon^{\mathbf{D}} \right|^{2} + \sqrt{2}\mathbf{K}_{*} \left| \varepsilon^{\mathbf{D}} \right| \;=\; \mathbf{W_1}(\varepsilon) + \mathbf{W_2}(\varepsilon) \tag{265}$$

Here $\mathbf{W_1}$ is the Newtonian potential of a viscous fluid and $\mathbf{W_2}$ is the plastic potential. By the Moreau–Rockafellar theorem we have

$$\partial \mathbf{W}(\varepsilon) = \partial \mathbf{W_1}(\varepsilon) + \partial \mathbf{W_2}(\varepsilon). \tag{266}$$

If $\left| \varepsilon^{D} \right| \neq 0$ then the relation reads

$$\partial \mathbf{W}(\varepsilon) = \mathbf{2}\mu\varepsilon^{\mathbf{D}} + \frac{\sqrt{2}\mathbf{K}_{*}}{\left| \varepsilon^{\mathbf{D}} \right|}\varepsilon^{\mathbf{D}}. \tag{267}$$

Let us define $\partial \mathbf{W}(\varepsilon)$ for the case $\left|\varepsilon^{\mathbf{D}}\right| = \mathbf{0}$. Evidently, $\mathbf{W_1}(\varepsilon) = \mathbf{0}$. To find $\partial \mathbf{W_2}$ we recall that by definition, $\boldsymbol{\tau} \in \partial W_2(\varepsilon)$ iff

$$\sqrt{\mathbf{2}}\mathbf{K}_* \left|æ^{\mathbf{D}}\right| - \sqrt{\mathbf{2}}\mathbf{K}_* \left|\varepsilon^{\mathbf{D}}\right| \geq \boldsymbol{\tau} : (æ - \varepsilon).$$

Note that

$$\boldsymbol{\tau} : (æ - \varepsilon) = \frac{\mathbf{1}}{\mathbf{n}} \mathbf{tr}\, \boldsymbol{\tau}\, (\mathbf{tr}\, æ - \mathbf{tr}\, \varepsilon) + \boldsymbol{\tau}^{\mathbf{D}} : æ^{\mathbf{D}},$$

Therefore,

$$\sqrt{\mathbf{2}}\mathbf{K}_* \left|æ^{\mathbf{D}}\right| \geq \frac{\mathbf{1}}{\mathbf{n}} \mathbf{tr}\, \boldsymbol{\tau}\, (\mathbf{tr}\, æ - \mathbf{tr}\, \varepsilon) + \boldsymbol{\tau}^{\mathbf{D}} : æ^{\mathbf{D}} \quad \forall æ^{\mathbf{D}}.$$

Let $\left| \mathbf{æ}^{\mathbf{D}} \right| = \mathbf{0}$, then

$$\mathbf{0} \geq \frac{\mathbf{1}}{\mathbf{n}} \mathbf{tr}\, \boldsymbol{\tau}\, (\mathbf{tr}\, \mathbf{æ} - \mathbf{tr}\, \varepsilon).$$

From here, it follows that

$$\mathbf{tr}\boldsymbol{\tau} = \mathbf{0}.$$

Therefore,

$$\sqrt{\mathbf{2}}\mathbf{K}_* \left| \mathbf{æ}^{\mathbf{D}} \right| \geq \boldsymbol{\tau}^{\mathbf{D}} : \mathbf{æ}^{\mathbf{D}} \quad \forall \mathbf{æ}^{\mathbf{D}}.$$

Without a loss of generality we may assume that $|\mathbf{æ}^{\mathbf{D}}| = \mathbf{1}$. To maximize the right–hand side we take $\mathbf{æ}^{\mathbf{D}} := \frac{\boldsymbol{\tau}^{\mathbf{D}}}{|\boldsymbol{\tau}^{\mathbf{D}}|}$. Then, we observe that

$$\sqrt{\mathbf{2}}\mathbf{K}_* \geq \left| \boldsymbol{\tau}^{\mathbf{D}} \right|.$$

Hence, $\partial \mathbf{W_2}(\varepsilon)$ for $\left|\varepsilon^{\mathbf{D}}\right| = \mathbf{0}$ consists of all $\boldsymbol{\tau}$ such that $\boldsymbol{\tau} = \boldsymbol{\tau}^D$ and

$$\left|\boldsymbol{\tau}^{\mathbf{D}}\right| \leq \sqrt{2}\mathbf{K}_*. \tag{268}$$

This relation reflects the physical nature of a **viscoplastic fluid**. In the stagnation zone the deviatoric part of its stress is not uniquely defined and can be any provides (268) is satisfied. Later we will see the examples exposing such a behavior of a solution.

Bingham flow in a basin

We consider a flow problem in a fixed domain, i.e., $\mathbf{\Omega_t} = \mathbf{\Omega}$ and the "no–slip" conditions $\mathbf{u} = \mathbf{u_0}$ are imposed on the boundary.
Classical solution is defined as $(\mathbf{u}, \boldsymbol{\sigma}, \mathbf{p})$ such that

$$\mathbf{u} \in \mathbf{C}(\overline{\mathbf{\Omega}}; \mathbb{R}^{\mathbf{d}}) \cap \mathbf{W}_2^{\mathbf{1}}(\mathbf{\Omega}; \mathbb{R}^{\mathbf{d}}), \tag{269}$$

$$\boldsymbol{\sigma} \in \mathbf{C}(\overline{\mathbf{\Omega}}; \mathbf{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}}) \cap \mathbf{W}_2^{\mathbf{1}}(\mathbf{\Omega}; \mathbf{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}}), \tag{270}$$

$$\mathbf{p} \in \mathbf{C}(\mathbf{\Omega}), \tag{271}$$

$$\frac{\partial \mathbf{u}}{\partial \mathbf{t}} \in \mathbf{L}_2(\mathbf{\Omega}; \mathbb{R}^{\mathbf{d}}), \tag{272}$$

$$\mathbf{f} \in \mathbf{L}_2(\mathbf{\Omega}; \mathbb{R}^{\mathbf{d}}), \tag{273}$$

Consider Dirichlét boundary conditions.

Let $\mathbf{v} \in \overset{\mathbf{o}}{\mathbf{H}}{}^{\mathbf{1}}(\boldsymbol{\Omega}; \mathbb{R}^{\mathbf{d}}) = \{ \mathbf{v} \in \mathbf{H}^{\mathbf{1}}(\boldsymbol{\Omega}; \mathbb{R}^{\mathbf{d}}) \mid \mathbf{v} = \mathbf{0} \text{ on } \partial\boldsymbol{\Omega} \}$. Multiply

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u_i} \frac{\partial \mathbf{u}}{\partial \mathbf{x_i}} - \mathbf{div}\ \boldsymbol{\sigma} = \mathbf{f}$$

by $\mathbf{v}$ and integrate by parts.

$$\int_{\boldsymbol{\Omega}} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{t}} \cdot \mathbf{v} + \mathbf{u_i} \frac{\partial \mathbf{u}}{\partial \mathbf{x_i}} \cdot \mathbf{v} - \mathbf{v}\, \mathbf{div}\, \boldsymbol{\sigma} \right) \mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{f v}\, \mathbf{dx},$$

$$\int_{\boldsymbol{\Omega}} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{t}} + \mathbf{u_i} \frac{\partial \mathbf{u}}{\partial \mathbf{x_i}} \right) \cdot \mathbf{v}\, \mathbf{dx} + \int_{\boldsymbol{\Omega}} \boldsymbol{\sigma} : \varepsilon(\mathbf{v})\, \mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{f v}\, \mathbf{dx}.$$

This relation may be rewritten as

$$\int_{\boldsymbol{\Omega}} \left(\frac{\partial \mathbf{u}}{\partial \mathbf{t}} + \mathbf{u_i} \frac{\partial \mathbf{u}}{\partial \mathbf{x_i}} \right) \cdot \mathbf{v}\, \mathbf{dx} + \int_{\boldsymbol{\Omega}} (\boldsymbol{\sigma} + \mathbf{p I}) : \varepsilon(\mathbf{v})\, \mathbf{dx} = \tag{274}$$

$$= \int_{\boldsymbol{\Omega}} (\mathbf{f v} + \mathbf{p}\, \mathbf{div}\, \mathbf{v})\, \mathbf{dx} \qquad \forall \mathbf{v} \in \overset{\mathbf{o}}{\mathbf{H}}{}^{\mathbf{1}}(\boldsymbol{\Omega}; \mathbb{R}^{\mathbf{d}})$$

By the constitutive relations,

$$\boldsymbol{\sigma} + \mathbf{pI} = \partial \mathbf{W_1} + \partial \mathbf{W_2} = \mathbf{2}\mu \varepsilon^{\mathbf{D}}(\mathbf{u}) + \partial \mathbf{W_2}(\varepsilon(\mathbf{u}))$$

what means that

$$\boldsymbol{\sigma} + \mathbf{pI} - \mathbf{2}\mu \varepsilon^{\mathbf{D}}(\mathbf{u}) \in \partial \mathbf{W_2}(\varepsilon(\mathbf{u}))$$

The latter means that

$$\sqrt{\mathbf{2}}\mathbf{k}_* \left| \text{æ}^{\mathbf{D}} \right| - \sqrt{\mathbf{2}}\mathbf{k}_* \left| \varepsilon^{\mathbf{D}}(\mathbf{u}) \right| \geq$$

$$\geq (\boldsymbol{\sigma} + \mathbf{pI} - \mathbf{2}\mu \varepsilon^{\mathbf{D}}(\mathbf{u})) : (\text{æ} - \varepsilon(\mathbf{u})) \quad \forall \text{æ} \in \mathbf{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}}$$

Set $\mathbf{æ} = \varepsilon(\mathbf{u} + \mathbf{v})$. Then,

$$\sqrt{2}\mathbf{k}_* \left| \mathbf{æ}^{\mathbf{D}} \right| - \sqrt{2}\mathbf{k}_* \left| \varepsilon^{\mathbf{D}}(\mathbf{u}) \right| \geq (\boldsymbol{\sigma} + \mathbf{pI} - 2\mu\varepsilon^{\mathbf{D}}(\mathbf{u})) : \varepsilon(\mathbf{v}).$$

and we obtain

$$\int_{\Omega} (\boldsymbol{\sigma} + \mathbf{pI}) : \varepsilon(\mathbf{v})\, \mathbf{dx} \leq$$

$$\leq \int_{\Omega} (2\mu\varepsilon^{\mathbf{D}}(\mathbf{u}) : \varepsilon^{\mathbf{D}}(\mathbf{v}) + \sqrt{2}\mathbf{k}_* \left| \mathbf{æ}^{\mathbf{D}} \right| - \sqrt{2}\mathbf{k}_* \left| \varepsilon^{\mathbf{D}}(\mathbf{u}) \right|) \mathbf{dx}$$

Now (274) yields inequality

$$\int_{\Omega} (\frac{\partial \mathbf{u}}{\partial \mathbf{t}} + \mathbf{u_i} \frac{\partial \mathbf{u}}{\partial \mathbf{x_i}}) \cdot \mathbf{v}\, \mathbf{dx} + \int_{\Omega} (2\mu\varepsilon^{\mathbf{D}}(\mathbf{u}) : \varepsilon^{\mathbf{D}}(\mathbf{v}) +$$

$$+ \sqrt{2}\mathbf{k}_* (\left| \mathbf{æ}^{\mathbf{D}} \right| - \left| \varepsilon^{\mathbf{D}}(\mathbf{u}) \right|) \mathbf{dx} \geq \int_{\Omega} (\mathbf{fv} + \mathbf{p}\, \mathbf{div}\, \mathbf{v}) \mathbf{dx} \; \forall \mathbf{v} \in \overset{\mathbf{o}}{\mathbf{H}}{}^{\mathbf{1}}(\Omega; \mathbb{R}^{\mathbf{d}})$$

If a trial function is taken from a narrower set

$$\mathbf{v} = \mathbf{w} + \mathbf{u_0} - \mathbf{u} \quad , \qquad \mathbf{w}, \mathbf{u_0} \in \mathring{\mathbf{J}}^{\mathbf{1}}_{\mathbf{2}}(\Omega)$$

so that

$$\mathbf{div}\, \mathbf{v} = \mathbf{0}, \quad \text{æ} = \varepsilon(\mathbf{w} + \mathbf{u_0})$$

Then (275) comes in a simpler form

$$\int_{\Omega} \Big\{ (\frac{\partial \mathbf{u}}{\partial \mathbf{t}} + \mathbf{u_i} \frac{\partial \mathbf{u}}{\partial \mathbf{x_i}}) \cdot (\mathbf{w} + \mathbf{u_0} - \mathbf{u}) + \mathbf{2}\mu \varepsilon^{\mathbf{D}}(\mathbf{u}) : \varepsilon^{\mathbf{D}}(\mathbf{w} + \mathbf{u_0} - \mathbf{u}) +$$

$$+ \sqrt{\mathbf{2}}\mathbf{k}_* \Big(\Big| \varepsilon^{\mathbf{D}}(\mathbf{w} + \mathbf{u_0}) \Big| - \Big| \varepsilon^{\mathbf{D}}(\mathbf{u}) \Big| \Big) \Big\} \mathbf{dx} \geq \int_{\Omega} \mathbf{f} \cdot (\mathbf{w} + \mathbf{u_0} - \mathbf{u}) \mathbf{dx} \; \forall \mathbf{w} \in \mathring{\mathbf{J}}^{\mathbf{1}}_{\mathbf{2}}(\Omega). \tag{275}$$

Denote $\tilde{\mathbf{u}} = \mathbf{u_0} + \mathbf{w}$, then we arrive at the variational inequality that describes the motion of an elastoplastic media

$$\int_\Omega \Big\{ \Big(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u_i}\frac{\partial \mathbf{u}}{\partial \mathbf{x_i}}\Big) \cdot (\tilde{\mathbf{u}} - \mathbf{u}) + \\ + 2\mu\varepsilon^{\mathbf{D}}(\mathbf{u}) : \varepsilon^{\mathbf{D}}(\tilde{\mathbf{u}} - \mathbf{u}) + \sqrt{2}\mathbf{k}_*\Big(\Big|\varepsilon^{\mathbf{D}}(\tilde{\mathbf{u}})\Big| - \Big|\varepsilon^{\mathbf{D}}(\mathbf{u})\Big|\Big)\Big\}\,\mathbf{dx} \geq \\ \geq \int_\Omega \mathbf{f}\cdot(\tilde{\mathbf{u}} - \mathbf{u})\,\mathbf{dx} \qquad \forall \tilde{\mathbf{u}} \in \mathbf{u_0} + \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\Omega) \quad (276)$$

Stationary flow

Stationary model is a particular case of the above that arises if $\partial \mathbf{u}/\partial \mathbf{t} = \mathbf{0}$. Here, we need to find $\mathbf{u} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}$ such that

$$\int_\Omega \Big\{ \mathbf{u_i} \frac{\partial \mathbf{u}}{\partial \mathbf{x_i}} \cdot (\mathbf{v} - \mathbf{u}) + \mathbf{2}\mu\varepsilon(\mathbf{u}) : \varepsilon(\mathbf{v} - \mathbf{u}) + \\ + \sqrt{\mathbf{2}}\mathbf{k}_*(|\,\varepsilon(\mathbf{v})\,| - |\,\varepsilon(\mathbf{u})\,|) \Big\}\, \mathbf{dx} \geq \int_\Omega \mathbf{f} \cdot (\mathbf{v} - \mathbf{u})\, \mathbf{dx} \;\forall \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}. \quad (277)$$

If, in addition, $\mathbf{v}$ is small (slow flows) then we arrive at the problem

$$\int_\Omega \big\{ \mathbf{2}\mu\varepsilon(\mathbf{u}) : \varepsilon(\mathbf{v} - \mathbf{u}) + \sqrt{\mathbf{2}}\mathbf{k}_*(|\,\varepsilon(\mathbf{v})\,| - |\,\varepsilon(\mathbf{u})\,|) \big\}\, \mathbf{dx} \geq \quad (278)$$

$$\geq \int_\Omega \mathbf{f} \cdot (\mathbf{v} - \mathbf{u})\, \mathbf{dx} \qquad \forall \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}$$

Existence and uniqueness

Consider the problem: find $\mathbf{u} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega}) + \mathbf{u_0}$ such that

$$\int_{\mathbf{\Omega}} \mathbf{2}\mu\varepsilon(\mathbf{u}) \cdot \varepsilon(\mathbf{v}-\mathbf{u}) + \sqrt{\mathbf{2}}\mathbf{K}_*(|\,\varepsilon(\mathbf{v})\,| - |\,\varepsilon(\mathbf{u})\,|)\mathbf{dx} \geq$$

$$\geq \int_{\mathbf{\Omega}} \mathbf{f} \cdot (\mathbf{v}-\mathbf{u})\mathbf{dx} \qquad \forall \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega}) + \mathbf{u_0}, \quad (279)$$

where $\mathbf{f} \in \mathbf{L_2}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}})$, $\mathbf{u_0} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega})$.
Uniqueness of the solution is easy to prove

Regularity of weak solutions

Regularity of weak solutions in the theory of Non Newtonian viscous fluids was investigated in the works of M. Fuchs, G. Seregin, M. Bildhauer and other authors. Readers can find a consequent exposition of these results in the book

M. Fuchs and G.A. Seregin. *Variational methods for problems from plasticity theory and for generalized Newtonian fluids.* Lect. Notes in Mathematics 1749, Springer-Verlag, Berlin (2000)

Let $\mathbf{u_1}$, $\mathbf{u_2} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega}) + \mathbf{u_0}$ be two different solutions. Then

$$\int_{\Omega} \{\mu\varepsilon(\mathbf{u_1}) \cdot \varepsilon(\mathbf{v} - \mathbf{u_1}) + \\ +\sqrt{\mathbf{2}}\mathbf{K}_*(|\,\varepsilon(\mathbf{v})\,| - |\,\varepsilon(\mathbf{u_1})\,|) - \mathbf{f} \cdot (\mathbf{v} - \mathbf{u_1})\}\mathbf{dx} \geq \mathbf{0}$$

$$\int_{\Omega} \{\mu\varepsilon(\mathbf{u_2}) \cdot \varepsilon(\mathbf{v} - \mathbf{u_2}) + \\ +\sqrt{\mathbf{2}}\mathbf{K}_*(|\,\varepsilon(\mathbf{v})\,| - |\,\varepsilon(\mathbf{u_2})\,|) - \mathbf{f} \cdot (\mathbf{v} - \mathbf{u_2})\}\mathbf{dx} \geq \mathbf{0}$$

Set $\mathbf{v} = \mathbf{u_2}$ in the first and $\mathbf{v} = \mathbf{u_1}$ in the second.

$$\int_{\Omega} \mu\varepsilon(\mathbf{u_1} - \mathbf{u_2}) \cdot \varepsilon(\mathbf{u_2} - \mathbf{u_1})\mathbf{dx} \geq \mathbf{0}$$

Thus,

$$\int_{\Omega} |\,\varepsilon(\mathbf{u_1} - \mathbf{u_2})\,|^{\mathbf{2}}\,\mathbf{dx} \leq \mathbf{0}$$

Existence

Existence follows from that the problem is equivalent to the variational problem

$$\mathbf{I(u) = \inf \{I(v) : v \in \overset{\circ}{J}{}_2^1(\Omega) + u_0\}}, \tag{280}$$

where

$$\mathbf{I(v)} = \int_{\mathbf{\Omega}} \{\mu \,|\, \varepsilon(\mathbf{u}) \,|^2 + \sqrt{\mathbf{2}}\mathbf{K}_* \,|\, \varepsilon(\mathbf{v}) \,| - \mathbf{f} \cdot \mathbf{v} \,\}\mathbf{dx}$$

This equivalence follow from the well known result (see, e.g., J.-L. Lions and G. Duvaut. Enequalities in mechanics and physics).

Theorem

Variational problem

$$\inf_{\mathbf{v}\in\mathbf{K}} \mathbf{J}(\mathbf{v}) = \mathbf{J}(\mathbf{v}_*); \qquad \mathbf{J}(\mathbf{v}) = \mathbf{J_0}(\mathbf{v}) + \mathbf{j}(\mathbf{v}), \tag{281}$$

where $\mathbf{K}$ *is a convex closed set,* $\mathbf{J_0} : \mathbf{V} \to \mathbb{R}$ *is a convex differentiable functional,* $\mathbf{j} : \mathbf{V} \to \mathbb{R}$ *is a convex functional,* $\mathbf{V}$ *has the same solution as the variational inequality*

$$(\mathbf{J_0'}(\mathbf{v}_*), \mathbf{v}_* - \mathbf{v}) + \mathbf{j}(\mathbf{v}_*) - \mathbf{j}(\mathbf{v}) \leq \mathbf{0}, \qquad \forall \mathbf{v} \in \mathbf{K} \tag{282}$$

In our case

$$\mathbf{J_0(v)} = \int_{\Omega} \mu \,|\, \varepsilon(\mathbf{u}) \,|^2 \,\mathbf{dx} - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \,\mathbf{dx}$$

$$\mathbf{j(v)} = \sqrt{2}\mathbf{K}_* \int_{\Omega} |\, \varepsilon(\mathbf{u}) \,|\, \mathbf{dx}$$

$$\mathbf{K} = \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\Omega) + \mathbf{u_0}, \qquad \mathbf{V} = \mathbf{H}^{\mathbf{1}}(\Omega, \mathbb{R}^{\mathbf{n}})$$

and we arrive at

$$\mathbf{I(v)} = \int_{\Omega} \{\mu \,|\, \varepsilon(\mathbf{u}) \,|^2 + \sqrt{2}\mathbf{K}_* \,|\, \varepsilon(\mathbf{v}) \,| - \mathbf{f} \cdot \mathbf{v} \,\}\mathbf{dx}$$

which is **convex, continuous, and coercive**. These properties imply existence.

Approximate solutions

Let $\mathbf{V_h} \subset \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega})$ and $\mathbf{dim\ V_h} < +\infty$, find $\mathbf{u_h} \in \mathbf{V_h} + \mathbf{u_0}$,

$$\int_{\mathbf{\Omega}} \mu\varepsilon(\mathbf{u_h}) \cdot \big(\varepsilon(\mathbf{v_h}) - \varepsilon(\mathbf{u_h})\big)\mathbf{dx} + \int_{\mathbf{\Omega}} \sqrt{\mathbf{2}}\mathbf{K}_* \big(|\,\varepsilon(\mathbf{v_h})\,| - |\,\varepsilon(\mathbf{u_h})\,|\big)\mathbf{dx} \geq \int_{\mathbf{\Omega}} \mathbf{f} \cdot (\mathbf{v_h} - \mathbf{u_h})\mathbf{dx}$$

$$\forall \mathbf{v_h} \in \mathbf{V_h} + \mathbf{u_0} \quad (283)$$

Variational formulation: find $\mathbf{u_h} \in \mathbf{V_h} + \mathbf{u_0}$ such that

$$\mathbf{I}(\mathbf{u_h}) = \mathbf{inf}\ \{\mathbf{I}(\mathbf{v_h}) :\ \mathbf{v_h} \in \mathbf{V_h} + \mathbf{u_0}\} \quad (284)$$

Serious difficulty is the condition

$$\mathbf{div\ v_h} = \mathbf{0}.$$

Example. Antiplane flow in a pipe

Consider a long tube of the cross-section Ω with the income and outcome pressure:

$$\mathbf{p}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{0}, \mathbf{t}) = \mathbf{0} \ , \ \mathbf{p}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{L}, \mathbf{t}) = -\mathbf{cL} \tag{285}$$

On the side surface

$$\mathbf{u} \mid_{\partial\Omega \times [\mathbf{0},\mathbf{L}]} = \mathbf{0}, \tag{286}$$

We are looking for a solution of the form

$$\mathbf{u}_1 = \mathbf{0}, \mathbf{u}_2 = \mathbf{0}, \mathbf{u}_3 = \mathbf{w}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{t})$$

$$\varepsilon(\mathbf{u}) = \begin{pmatrix} 0 & 0 & \frac{1}{2}\mathbf{w}_{,1} \\ 0 & 0 & \frac{1}{2}\mathbf{w}_{,2} \\ \frac{1}{2}\mathbf{w}_{,1} & \frac{1}{2}\mathbf{w}_{,2} & 0 \end{pmatrix}$$

Since

$$\varepsilon(\mathbf{u}) = \begin{pmatrix} 0 & 0 & \frac{1}{2}\mathbf{w}_{,1} \\ 0 & 0 & \frac{1}{2}\mathbf{w}_{,2} \\ \frac{1}{2}\mathbf{w}_{,1} & \frac{1}{2}\mathbf{w}_{,2} & 0 \end{pmatrix}$$

we have

$$|\,\varepsilon(\mathbf{u})\,|^2 = \frac{\mathbf{1}}{\mathbf{2}}\,|\,\nabla \mathbf{w}\,|^2\,,\ \frac{\partial \mathbf{u}}{\partial \mathbf{x_1}} = (\mathbf{0}, \mathbf{0}, \mathbf{w}_{,1}),$$

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x_2}} = (\mathbf{0}, \mathbf{0}, \mathbf{w}_{,2}), \quad \frac{\partial \mathbf{u}}{\partial \mathbf{x_3}} = (\mathbf{0}, \mathbf{0}, \mathbf{0}), \quad \mathbf{u_i}\frac{\partial \mathbf{u}}{\partial \mathbf{x_i}} = \mathbf{0}$$

$$\mathbf{u_i}\frac{\partial \mathbf{u}}{\partial \mathbf{x_i}} = \mathbf{0}\ ,\ \mathbf{divu} = \mathbf{0} \tag{287}$$

Remark

If set in (277) $\mathbf{v} = (\mathbf{0}, \mathbf{0}, \tilde{\mathbf{w}})$, where $\tilde{\mathbf{w}} \in \overset{\mathbf{o}}{\mathbf{H}}{}^{\mathbf{1}}(\boldsymbol{\Omega})$, $\tilde{\mathbf{w}} = \tilde{\mathbf{w}}(x_1, x_2)$, $\mathbf{u_0} = \mathbf{0}$ and $\mathbf{f} = (\mathbf{0}, \mathbf{0}, \mathbf{c})$, then we arrive at the problem

$$\int_{\boldsymbol{\Omega}} \left\{ \mu \nabla \mathbf{w} \cdot (\nabla \tilde{\mathbf{w}} - \nabla \mathbf{w}) + \mathbf{k}_* (|\, \nabla \tilde{\mathbf{w}} \,| - |\, \nabla \mathbf{w} \,|) \right\} \mathbf{dx} \geq$$

$$\geq \int_{\boldsymbol{\Omega}} \mathbf{c}(\tilde{\mathbf{w}} - \mathbf{w})\, \mathbf{dx} \qquad \forall \tilde{\mathbf{w}} \in \overset{\mathbf{o}}{\mathbf{H}}{}^{\mathbf{1}}(\boldsymbol{\Omega}).$$

We see that stationary slow flow of a viscoplastic fluid in a pipe is described by a **variational inequality**.

In this case,

$$\boldsymbol{\sigma} + \mathbf{p}\mathbf{I} = \boldsymbol{\tau} = \begin{pmatrix} 0 & 0 & \tau_{31}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{t}) \\ 0 & 0 & \tau_{32}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{t}) \\ \tau_{31}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{t}) & \tau_{32}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{t}) & 0 \end{pmatrix}$$

and $\boldsymbol{\tau} = \boldsymbol{\tau}^D$.
Let $\rho_0 = 1$. Take the third equation of the motion:

$$\frac{\partial \mathbf{w}}{\partial \mathbf{t}} = \boldsymbol{\sigma}_{31,1} + \boldsymbol{\sigma}_{32,2} + \boldsymbol{\sigma}_{33,3}\,, \tag{288}$$

Here $\boldsymbol{\sigma} = \boldsymbol{\tau} - p\mathbf{I}$ and $\boldsymbol{\sigma}_{33} = -p$.

Therefore, we have

$$\frac{\partial \mathbf{w}}{\partial \mathbf{t}} = \boldsymbol{\sigma}_{31,1} + \boldsymbol{\sigma}_{32,2} - \frac{\partial \mathbf{p}}{\partial \mathbf{x_3}} \tag{289}$$

Since $\boldsymbol{\sigma}_{31}, \boldsymbol{\sigma}_{32}$, and $\mathbf{w}$ depend only on x_1, x_2, we rewrite 289 as follows:

$$\frac{\partial \mathbf{p}}{\partial \mathbf{x_3}} = \mathbf{a}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{t})$$

Thus,

$$\mathbf{p}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \mathbf{t}) = \mathbf{a}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{t}) + \mathbf{x_3}\mathbf{b}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{t})\,.$$

Boundary conditions (285) say that

$$\mathbf{p}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{0}, \mathbf{t}) = \mathbf{a}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{t}) = \mathbf{0}\,,$$
$$\mathbf{p}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{L}, \mathbf{t}) = \mathbf{L}\mathbf{b}(\mathbf{x_1}, \mathbf{x_2}, \mathbf{t}) = -\mathbf{c}\mathbf{L} \implies \mathbf{b} = -\mathbf{c}\,.$$

Hence, the pressure is found as $\mathbf{p} = -\mathbf{C}\mathbf{x_3}$.

Now the motion is governed by the single equation

$$\frac{\partial \mathbf{w}}{\partial \mathbf{t}} = \frac{\partial \boldsymbol{\sigma}_{31}}{\partial \mathbf{x}_1} + \frac{\partial \boldsymbol{\sigma}_{32}}{\partial \mathbf{x}_2} + \mathbf{c} \tag{290}$$

Recall that

$$| \varepsilon(\mathbf{u}) |^2 = \frac{\mathbf{1}}{\mathbf{2}} | \nabla \mathbf{w} |^2$$

Rewrite the constitutive relation

$$\boldsymbol{\sigma} + \mathbf{pI} = \begin{cases} 2\mu\varepsilon(\mathbf{u}) + \sqrt{\mathbf{2}}\mathbf{K}_* \dfrac{\varepsilon(\mathbf{u})}{| \varepsilon(\mathbf{u}) |} & \text{for } | \varepsilon(\mathbf{u}) | > \mathbf{0} \\ \boldsymbol{\tau} = \boldsymbol{\tau}^{\mathbf{D}}, | \boldsymbol{\tau}^{\mathbf{D}} | \leq \sqrt{\mathbf{2}}\mathbf{K}_* & \text{for } | \varepsilon(\mathbf{u}) | = \mathbf{0} \end{cases}$$

in terms of the problem considered.

We have

$$\boldsymbol{\sigma}_{3\mathbf{i}} = \begin{cases} (2\mu + \sqrt{2}K_* \dfrac{1}{\dfrac{1}{\sqrt{2}} \mid \nabla \mathbf{w} \mid}) \dfrac{1}{2} \mathbf{w}_{,\mathbf{i}} & \mid \nabla \mathbf{w} \mid > \mathbf{0}\,, \\ & \mathbf{i} = \mathbf{1}, \mathbf{2}\,. \\ \boldsymbol{\tau}_{3\mathbf{i}}, \quad \mid \boldsymbol{\tau}^{\mathbf{D}} \mid \leq \sqrt{\mathbf{2}} \mathbf{K}_* & \mid \nabla \mathbf{w} \mid = \mathbf{0}\,, \end{cases} \tag{291}$$

where $\mid \boldsymbol{\tau}^{\mathbf{D}} \mid^2 = \mathbf{2}\boldsymbol{\tau}_{\mathbf{31}}{}^2 + \mathbf{2}\boldsymbol{\tau}_{\mathbf{32}}{}^2$. Therefore,

$$\boldsymbol{\sigma}_{3\mathbf{i}} = \begin{cases} (\mu + \dfrac{\mathbf{K}_*}{\mid \nabla \mathbf{w} \mid}) \mathbf{w}_{,\mathbf{i}}, & \text{for} \quad \mid \nabla \mathbf{w} \mid > \mathbf{0}\,, \\ \boldsymbol{\tau}_{3\mathbf{i}}, \quad \sqrt{\boldsymbol{\tau}_{\mathbf{31}}{}^2 + \boldsymbol{\tau}_{\mathbf{32}}{}^2} \leq \mathbf{K}_*, & \text{for} \quad \mid \nabla \mathbf{w} \mid = \mathbf{0}\,. \end{cases} \tag{292}$$

We observe that

$$| \boldsymbol{\sigma_3} | := \sqrt{(\boldsymbol{\sigma_{31}^2} + \boldsymbol{\sigma_{32}^2})} = \begin{cases} \mu \mid \nabla \mathbf{w} \mid + \mathbf{K}_*, & \text{for } \mid \nabla \mathbf{w} \mid > \mathbf{0}\,, \\ \sqrt{\boldsymbol{\tau_{31}^2} + \boldsymbol{\tau_{32}^2}} \leq \mathbf{K}_*, & \text{for } \mid \nabla \mathbf{w} \mid = \mathbf{0} \end{cases} .$$

Now (292), the equation

$$\frac{\partial \mathbf{w}}{\partial \mathbf{t}} = \mathbf{div} \boldsymbol{\sigma_3} + \mathbf{c}\,.$$

and the condition

$$\mathbf{w} \mid_{\partial \Omega} = \mathbf{0} \tag{293}$$

describes the solution.

Cylindrical pipe

Consider stationary flow in a cylindrical pipe.

$$\boldsymbol{\sigma_3} := \eta = (\eta_\rho, \eta_\varphi)\,, \qquad \eta_\rho = \boldsymbol{\sigma_{3\rho}}, \qquad \eta_\varphi = \boldsymbol{\sigma_{3\varphi}}\,.$$

Problem is axisymmetric and

$$\boldsymbol{\sigma}_{3\varphi} = 0, \quad \frac{\partial \mathbf{w}}{\partial \varphi} = 0\,.$$

Problem is stationary:

$$\frac{\partial \mathbf{w}}{\partial t} = 0\,.$$

In the axisymmetric case

$$\mathbf{div}(\boldsymbol{\sigma_3}) = \mathbf{div}\eta = \frac{\partial \eta_\rho}{\partial \rho} + \frac{\eta_\rho}{\rho}\,,$$

Now, the motion equation reads

$$\frac{\partial \eta_\rho}{\partial \rho} + \frac{\eta_\rho}{\rho} + \mathbf{c} = \mathbf{0} \implies \frac{\mathbf{1}}{\rho}\frac{\partial(\rho\eta_\rho)}{\partial \rho} + \mathbf{c} = \mathbf{0} \tag{294}$$

$$\eta_\rho = \begin{cases} (\mu + \dfrac{\mathbf{K}_*}{|\dfrac{\partial \mathbf{w}}{\partial \rho}|})\dfrac{\partial \mathbf{w}}{\partial \rho} & \text{for } |\dfrac{\partial \mathbf{w}}{\partial \rho}| > \mathbf{0}\,, \\ \tilde{\eta_\rho}, |\tilde{\eta_\rho}| \leq \mathbf{K}_* & \text{for } |\dfrac{\partial \mathbf{w}}{\partial \rho}| = 0 \end{cases} . \tag{295}$$

From the viewpoint of physics it would be natural to await the following behavior of the media:

1. If the "pressure grade" c is small, then there is no motion at all: plastic properties of the media dominate and does not allow any motion.
2. If c becomes large enough, then the motion starts in places where the effective stresses achieve the critical value, i.e. near the boundary. Central part moves as a solid body.

Formal analysis confirms these expectations. Integrate (294), then we observe that

$$\eta_\rho = -\frac{c}{2}\rho + \frac{C_2}{\rho} \tag{296}$$

$\sigma_{3\rho}$ at the center must be finite, therefore $C_2 = 0$ and (296) reads

$$\eta_\rho = -\frac{c}{2}\rho \tag{297}$$

For which c no motion arise?

$$| \eta_\rho(\rho) | \leq | \eta_\rho(R) | = \frac{c}{2}R \leq K_*,$$

Hence, these values are: $c \leq \frac{2K_*}{R}$.

Indeed, in such a case we have a function η_ρ that satisfies the equation and constitutive relations related to the branch $\mid \dfrac{\partial \mathbf{w}}{\partial \rho} \mid = 0$. Since $\mathbf{w} \mid_{\partial\Omega} = 0$ we conclude that $\mathbf{w} \equiv \mathbf{0}$.

Let $\mathbf{c} > \dfrac{\mathbf{2K}_*}{\mathbf{R}}$. We are looking for a solution such that

1. for $R_* \leq \rho \leq R$ the media is deformed and $\mid \dfrac{\partial \mathbf{w}}{\partial \rho} \mid > 0$
2. for $0 \leq \rho \leq R_*$ the media is rigid and $\mid \dfrac{\partial \mathbf{w}}{\partial \rho} \mid = 0$.

Then $\mathbf{R}_*$ is defined by the relation

$$| \eta_\rho(\mathbf{R}_*) | = \frac{\mathbf{c}}{\mathbf{2}}\mathbf{R}_* = \mathbf{K}_* \Longrightarrow \mathbf{R}_* = \frac{\mathbf{2K}_*}{\mathbf{c}} < \mathbf{R}$$

For $\mathbf{R}_* \le \rho \le \mathbf{R}$ it should be $\dfrac{\partial \mathbf{w}}{\partial \rho} < 0$, so that

$$\eta_\rho \;=\; \left(\mu + \frac{\mathbf{K}_*}{| \dfrac{\partial \mathbf{w}}{\partial \rho} |} \right) \frac{\mathbf{dw}}{\mathbf{d}\rho} \;=\; \mu \frac{\mathbf{dw}}{\mathbf{d}\rho} - \mathbf{K}_* \,.$$

The latter relation leads to the conclusion that

$$\mu \frac{\mathbf{dw}}{\mathbf{d}\rho} - \mathbf{K}_* = -\frac{\mathbf{c}}{\mathbf{2}}\rho \Longrightarrow \quad \mu \mathbf{w} - \mathbf{K}_*\rho = -\frac{\mathbf{c}}{\mathbf{4}}\rho^2 + \mathbf{C_3}$$

$$\mathbf{0} = \mathbf{K}_*\mathbf{R} - \frac{\mathbf{c}}{\mathbf{4}}\mathbf{R}^2 + \mathbf{C_3} \Longrightarrow \quad \mathbf{C_3} = \frac{\mathbf{c}}{\mathbf{4}}\mathbf{R}^2 - \mathbf{K}_*\mathbf{R}\,. \qquad (298)$$

Consequently $\mu \mathbf{w} = \mathbf{K}_*(\rho - \mathbf{R}) - \dfrac{\mathbf{c}}{\mathbf{4}}(\rho^2 - \mathbf{R}^2)$.

Therefore, we arrive at the conclusion that for $\mathbf{c} > \dfrac{\mathbf{2K_*}}{\mathbf{R}}$ the solution is as follows:

$$\mathbf{w} = \begin{cases} \dfrac{\mathbf{1}}{\mu}\left[(\mathbf{R_*} - \mathbf{R}) - \dfrac{\mathbf{c}}{\mathbf{4}}(\mathbf{R_*}^{\mathbf{2}} - \mathbf{R^2})\right] & \mathbf{0} \le \rho \le \mathbf{R_*}\,, \\ \dfrac{\mathbf{1}}{\mu}\left[(\rho - \mathbf{R}) - \dfrac{\mathbf{c}}{\mathbf{4}}(\rho^{\mathbf{2}} - \mathbf{R^2})\right] & \mathbf{R_*} \le \rho \le \mathbf{R}\,. \end{cases}$$

Thus,

$$\frac{\mathbf{dw}}{\mathbf{d}\rho} = \mathbf{0} \quad \text{for} \quad \mathbf{0} \le \rho \le \mathbf{R_*}$$

and

$$\frac{\mathbf{1}}{\mu}\left[\mathbf{1} - \frac{\mathbf{c}}{\mathbf{2}}\rho\right] \quad \text{for} \quad \mathbf{R_*} \le \rho \le \mathbf{R}\,.$$

FUNCTIONAL A POSTERIORI ESTIMATES FOR GENERALIZED NEWTONIAN FLUIDS

For the considered class of problems it is convenient to derive estimates of deviations from exact solutions by the variational method. First, we establish the estimate

$$\frac{\nu}{2}\int_{\Omega} |\,\varepsilon(\mathbf{v}-\mathbf{u})\,|^2\,\mathbf{dx} \le \mathbf{J}(\mathbf{v})-\mathbf{J}(\mathbf{u}), \tag{299}$$

where $\mathbf{v}$ is an arbitrary function in $\overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\Omega)+\mathbf{u_0}$. Indeed,

$$\mathbf{J}(\mathbf{v})-\mathbf{J}(\mathbf{u}) = \int_{\Omega}\Big(\frac{\nu}{2}\,|\,\varepsilon(\mathbf{v}-\mathbf{u})\,|^2+\nu\varepsilon(\mathbf{u}):\varepsilon(\mathbf{v}-\mathbf{u})+$$

$$+\sqrt{2}\mathbf{K}_*(|\varepsilon(\mathbf{v})|-|\varepsilon(\mathbf{u})|)\Big)\mathbf{dx}-\int_{\Omega}\mathbf{f}\cdot(\mathbf{v}-\mathbf{u})\,\mathbf{dx}.$$

To estimate J(u) from below, we construct a set of variational problems whose functionals are defined on the functional class wider than $\overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}$. These problems we shall call "disturbed".

Let us define two functions $\mathbf{q} \in \overset{\circ}{\mathbf{L}}{}_{\mathbf{2}}(\boldsymbol{\Omega})$ and $\boldsymbol{\tau}_2 \in \boldsymbol{\Sigma}$ and the functional

$$\bar{\mathbf{J}}(\mathbf{v}) := \int_{\boldsymbol{\Omega}} \left(\frac{\nu}{\mathbf{2}} \, | \, \boldsymbol{\varepsilon}(\mathbf{v}) \, |^{\mathbf{2}} + \boldsymbol{\tau}_{\mathbf{2}} : \boldsymbol{\varepsilon}(\mathbf{v}) - \psi^*(\boldsymbol{\tau}_{\mathbf{2}}) - \mathbf{f} \cdot \mathbf{v} - \mathbf{q} \, \mathrm{div}(\mathbf{v} - \mathbf{u_0}) \right) \mathbf{dx},$$

where $\psi^* : \mathbb{M}^{\mathbf{d} \times \mathbf{d}} \to \mathbb{R}$ is the functional conjugate to $\psi(\kappa) := \sqrt{2} K_* |\kappa|$ in the sense of Young–Fenchel, i.e.,

$$\psi^*(\kappa^*) = \sup_{\kappa \in \mathbb{M}^{\mathbf{d} \times \mathbf{d}}} \left\{ \kappa^* : \kappa - \psi(\kappa) \right\}.$$

Now, the following variational problem $\bar{\mathcal{P}}_{q,\tau_2}$ arises: Find $\bar{\mathbf{u}} \in \mathbf{V_0} + \mathbf{u_0}$ such that

$$\bar{\mathbf{J}}(\bar{\mathbf{u}}) = \inf \bar{\mathcal{P}} := \inf_{\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}} \bar{\mathbf{J}}(\mathbf{v}) \, .$$

Here, in principle, $\bar{\mathbf{u}} = \bar{\mathbf{u}}_{\mathbf{q},\tau_2}$. However, for the sake of simplicity we shall not do this assuming that the bar above means that a quantity depends on the above functions.

Problem $\bar{\mathcal{P}}$ is uniquely solvable and

$$\inf \bar{\mathcal{P}} \leq \inf \mathcal{P}. \tag{300}$$

Existence and uniqueness of $\bar{\mathbf{u}}$ follows from the properties of the convex functional $\bar{\mathbf{J}}$ and closed set $\mathbf{V_0} + \mathbf{u_0}$. In accordance with the definition of ψ^*, for any $\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\boldsymbol{\Omega}) + \mathbf{u_0}$ we have the inequality

$$\bar{\mathbf{J}}(\mathbf{v}) = \int_{\Omega} \left(\frac{\nu}{2} \, | \, \boldsymbol{\varepsilon}(\mathbf{v}) \, |^2 + \boldsymbol{\tau}_2 : \boldsymbol{\varepsilon}(\mathbf{v}) - \psi^*(\boldsymbol{\tau}_2) \right) \mathbf{dx} - \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \mathbf{dx} \leq \mathbf{J}(\mathbf{v}).$$

Therefore,

$$\inf_{\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}} \bar{\mathbf{J}}(\mathbf{v}) \leq \inf_{\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\boldsymbol{\Omega}) + \mathbf{u_0}} \bar{\mathbf{J}}(\mathbf{v}) \leq \inf_{\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\boldsymbol{\Omega}) + \mathbf{u_0}} \mathbf{J}(\mathbf{v}) = \inf \mathcal{P}.$$

Now, we observe that

$$\frac{\nu}{2}\int_{\Omega} |\,\varepsilon(\mathbf{v}-\mathbf{u})\,|^2\,\mathbf{dx} \le \mathbf{J}(\mathbf{v}) - \inf \bar{\mathcal{P}}. \tag{301}$$

However, the value of $\inf \bar{\mathcal{P}}$ is unknown! To overcome this difficulty we attract dual variational problem $\bar{\mathcal{P}}^*$.
If $\inf \bar{\mathcal{P}} = \sup \bar{\mathcal{P}}^*$, then $\inf \bar{\mathcal{P}}$ can be replaced by a lower estimate of $\sup \bar{\mathcal{P}}^*$. Estimates obtained on this way will depend on the functions $\boldsymbol{\tau}_2$ and $\mathbf{q}$ and also on the variables of the dual problem. Note that there are different variational problem that may be viewed as dual to $\bar{\mathcal{P}}$. The problem is to find in this collection a proper variant that leads to estimates convenient for practice and having a good accuracy. For the class of problems considered, the following variant is possible.

Define the Lagrangian

$$\bar{\mathsf{L}}(\mathbf{v};\boldsymbol{\tau}_1) := \int_{\Omega} \Big(\boldsymbol{\varepsilon}(\mathbf{v}) : (\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2) - \frac{1}{2\nu} \mid \boldsymbol{\tau}_1 \mid^2 - \psi^*(\boldsymbol{\tau}_2) \Big) \mathbf{dx}$$
$$- \int_{\Omega} \mathbf{f} \cdot \mathbf{v} \mathbf{dx} - \int_{\Omega} \mathbf{q} \cdot \mathrm{div} \cdot (\mathbf{v} - \mathbf{u_0}) \mathbf{dx}.$$

Then,

$$\bar{\mathsf{J}}(\mathbf{v}) = \sup_{\boldsymbol{\tau}_1 \in \boldsymbol{\Sigma}} \bar{\mathsf{L}}(\mathbf{v};\boldsymbol{\tau}_1),$$

so that Problem $\bar{\mathcal{P}}$ is equivalent to the minimax problem $\inf_{\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}} \sup_{\boldsymbol{\tau}_1 \in \boldsymbol{\Sigma}} \bar{\mathsf{L}}(\mathbf{v};\boldsymbol{\tau}_1)$. The respective dual problem is

$$\sup_{\boldsymbol{\tau}_1 \in \boldsymbol{\Sigma}} \inf_{\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}} \bar{\mathsf{L}}(\mathbf{v};\boldsymbol{\tau}_1).$$

Note that

$$\inf_{\mathbf{v}\in\mathbf{V_0}+\mathbf{u_0}} \bar{\mathsf{L}}(\mathbf{v};\tau_1) = \begin{cases} \bar{\mathsf{I}}(\tau_1) & \text{if } \tau_1 \in \bar{\mathbf{\Sigma}}_{\mathbf{f}}(\Omega), \\ -\infty & \text{if } \tau_1 \notin \bar{\mathbf{\Sigma}}_{\mathbf{f}}(\Omega), \end{cases}$$

where

$$\bar{\mathsf{I}}(\tau_1) = \int_{\Omega} \Big(\varepsilon(\mathbf{u_0}) : (\tau_1 + \tau_2) - \frac{1}{2\nu} \mid \tau_1 \mid^2 - \psi^*(\tau_2) - \mathbf{f}\cdot\mathbf{u_0} \Big) \mathbf{dx},$$

and $\bar{\mathbf{\Sigma}}_{\mathbf{f}}$ is an affine subset in $\mathbf{\Sigma}$ that consists of the functions τ satisfying (in a generalized sense) the condition $\mathbf{div}(\tau + \tau_2) = \nabla\mathbf{q} - \mathbf{f}$, i.e.,

$$\bar{\mathbf{\Sigma}}_{\mathbf{f}}(\Omega) := \Big\{ \tau \in \mathbf{\Sigma}(\Omega) \;\Big|\; \int_{\Omega} \varepsilon(\mathbf{w}) : (\tau + \tau_2)\, \mathbf{dx} = \int_{\Omega} (\mathbf{f}\cdot\mathbf{w} + \mathbf{q}\,\mathrm{div}\,\mathbf{w})\, \mathbf{dx},\ \mathbf{w} \in \mathbf{V_0} \Big\}.$$

Thus, we arrive at the following formulation of the Dual Variational Problem.

Problem $\bar{\mathcal{P}}^*$ For given $\boldsymbol{\tau}_2 \in \boldsymbol{\Sigma}$ and $\mathbf{q} \in \mathring{\mathbf{L}}_2(\Omega)$ find a function $\overline{\boldsymbol{\sigma}}_1 \in \bar{\boldsymbol{\Sigma}}_{\mathbf{f}}(\Omega)$ such that

$$\bar{\mathsf{I}}(\overline{\boldsymbol{\sigma}}_1) = \sup \bar{\mathcal{P}}^* := \sup_{\boldsymbol{\tau}_1 \in \bar{\boldsymbol{\Sigma}}_{\mathbf{f}}} \bar{\mathsf{I}}(\boldsymbol{\tau}_1)\,.$$

Theorem

Problem $\bar{\mathcal{P}}^*$ *has a unique solution* $\overline{\sigma}_1$ *satisfying the conditions*

$$\inf \bar{\mathcal{P}} = \bar{\mathbf{J}}(\bar{\mathbf{u}}) \;=\; \sup \bar{\mathcal{P}}^* \;= \bar{\mathbf{I}}(\overline{\sigma}_1)\,, \tag{302}$$

$$\nu\varepsilon(\bar{\mathbf{u}}) = \overline{\sigma}_1. \tag{303}$$

Proof.

$-\bar{\mathsf{I}}$ is strictly convex and $\bar{\mathbf{\Sigma}}_{\mathbf{f}}$ is a convex and closed subset of $\mathbf{\Sigma}$. Therefore, Problem $\bar{\mathcal{P}}^*$ has a unique solution.
Note that $\sup \bar{\mathcal{P}}^* \leq \inf \bar{\mathcal{P}}$. This fact follows from the relation

$$\sup \inf \bar{\mathsf{L}} \leq \inf \sup \bar{\mathsf{L}}$$

$\bar{\mathbf{u}}$ satisfies the integral identity

$$\int_{\Omega} (\nu \varepsilon(\bar{\mathbf{u}}) : \varepsilon(\mathbf{w}) + \boldsymbol{\tau}_2 : \varepsilon(\mathbf{w}))\, \mathbf{dx} = \int_{\Omega} (\mathbf{f} \cdot \mathbf{w} + \mathbf{q}\mathrm{div}\mathbf{w})\mathbf{dx}\ \mathbf{w} \in \mathbf{V_0}.$$

From here, it follows that

$$\int_{\Omega} \mathbf{f} \cdot \bar{\mathbf{u}}\mathbf{dx} = \int_{\Omega} (\nu \,|\, \varepsilon(\bar{\mathbf{u}}) \,|^2 - \nu \varepsilon(\bar{\mathbf{u}}) : \varepsilon(\mathbf{u_0}) +$$
$$+ \boldsymbol{\tau}_2 : \varepsilon(\bar{\mathbf{u}} - \mathbf{u_0}) + \mathbf{f} \cdot \mathbf{u_0} - \mathbf{q}\mathrm{div}(\bar{\mathbf{u}} - \mathbf{u_0}))\mathbf{dx}.$$

Therefore,

$$\begin{aligned}
&\inf \bar{\mathcal{P}} = \bar{\mathsf{J}}(\bar{\mathbf{u}}) \\
&= \int_{\Omega} \left(\frac{\nu}{2} \left| \varepsilon(\bar{\mathbf{u}}) \right|^2 + \boldsymbol{\tau}_2 : \varepsilon(\bar{\mathbf{u}}) - \psi^*(\boldsymbol{\tau}_2) \right) \mathbf{dx} - \int_{\Omega} (\mathbf{q} \mathrm{div}(\bar{\mathbf{u}} - \mathbf{u_0}) + \mathbf{f} \cdot \bar{\mathbf{u}}) \mathbf{dx} \\
&= \int_{\Omega} \left(\boldsymbol{\tau}_2 : \varepsilon(\mathbf{u_0}) - \frac{\nu}{2} \left| \varepsilon(\bar{\mathbf{u}}) \right|^2 - \psi^*(\boldsymbol{\tau}_2) + \nu \varepsilon(\mathbf{u_0}) : \varepsilon(\bar{\mathbf{u}}) \right) \mathbf{dx} - \int_{\Omega} \mathbf{f} \cdot \mathbf{u_0} \mathbf{dx}.
\end{aligned}$$

Since $\nu\varepsilon(\bar{\mathbf{u}}) \in \bar{\boldsymbol{\Sigma}}_{\mathbf{f}}$, we know that

$$\bar{\mathsf{I}}(\nu\varepsilon(\bar{\mathbf{u}})) \leq \sup \bar{\mathcal{P}}^*$$

$$\bar{\mathsf{I}}(\nu\varepsilon(\bar{\mathbf{u}})) =$$

$$= \int_\Omega \left(\varepsilon(\mathbf{u_0}) : (\nu\varepsilon(\bar{\mathbf{u}}) + \boldsymbol{\tau_2}) - \frac{\nu}{2} \mid \varepsilon(\bar{\mathbf{u}}) \mid^2 - \psi^*(\boldsymbol{\tau_2})\right) \mathbf{dx} - \int_\Omega \mathbf{f} \cdot \mathbf{u_0} \mathbf{dx} \leq$$

$$\leq \sup \bar{\boldsymbol{\mathcal{P}}}^* \leq \inf \bar{\boldsymbol{\mathcal{P}}} =$$

$$= \int_\Omega \left(\boldsymbol{\tau_2} : \varepsilon(\mathbf{u_0}) - \frac{\nu}{2} \mid \varepsilon(\bar{\mathbf{u}}) \mid^2 - \psi^*(\boldsymbol{\tau_2}) + \nu\varepsilon(\mathbf{u_0}) : \varepsilon(\bar{\mathbf{u}})\right) \mathbf{dx} - \int_\Omega \mathbf{f} \cdot \mathbf{u_0}\, \mathbf{dx}.$$

Consequently, $\nu\varepsilon(\bar{\mathbf{u}}) = \overline{\boldsymbol{\sigma}}_1$ and

$$\inf \bar{\boldsymbol{\mathcal{P}}} = \sup \bar{\boldsymbol{\mathcal{P}}}^*.$$

Estimates of deviations from exact solutions for solenoidal fields

Let $\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}$. For it and $\boldsymbol{\tau}_{1f} \in \bar{\boldsymbol{\Sigma}}_{\mathbf{f}}(\Omega)$ the inequality

$$\int_{\Omega} \frac{\nu}{2} \mid \varepsilon(\mathbf{v}-\mathbf{u}) \mid^2 \mathbf{dx} \leq \tag{304}$$

$$\leq \mathbf{J}(\mathbf{v}) - \inf \bar{\mathcal{P}} = \mathbf{J}(\mathbf{v}) - \sup \bar{\mathcal{P}}^* \leq \mathbf{J}(\mathbf{v}) - \bar{\mathbf{I}}(\boldsymbol{\tau}_{1\mathbf{f}})$$

holds. Estimate the right–hand side of (304) as follows:

$$\mathbf{J}(\bar{\mathbf{v}}) - \bar{\mathbf{I}}(\boldsymbol{\tau}_{1\mathbf{f}}) \leq \int_{\Omega} \left(\frac{\nu}{2} \mid \varepsilon(\mathbf{v}) \mid^2 + \frac{\mathbf{1}}{2\nu} \mid \boldsymbol{\tau}_{1\mathbf{f}} \mid^2 - \varepsilon(\mathbf{u_0}) : \boldsymbol{\tau}_{1\mathbf{f}} \right) \mathbf{dx}$$

$$+ \int_{\Omega} (\psi(\varepsilon(\mathbf{v})) + \psi^*(\boldsymbol{\tau}_2) - \varepsilon(\mathbf{u_0}) : \boldsymbol{\tau}_2) \, \mathbf{dx} + \int_{\Omega} \mathbf{f} \cdot (\mathbf{u_0} - \mathbf{v}) \mathbf{dx} \, .$$

Let $\mathbf{q} \in \mathring{\mathbf{L}}_2(\Omega)$. Since $\boldsymbol{\tau}_{1f} \in \bar{\boldsymbol{\Sigma}}_{\mathbf{f}}(\Omega)$ and $\mathbf{v} \in \mathring{\mathbf{J}}^1_2(\Omega) + \mathbf{u_0}$, we have

$$\int_\Omega \mathbf{f} \cdot (\mathbf{u_0} - \mathbf{v})\, \mathbf{dx} = \int_\Omega (\mathbf{f} \cdot (\mathbf{u_0} - \mathbf{v}) + \mathbf{q}\mathrm{div}(\mathbf{u_0} - \mathbf{v}))\mathbf{dx}$$

$$= \int_\Omega \varepsilon(\mathbf{u_0} - \mathbf{v}) : (\boldsymbol{\tau}_{1\mathbf{f}} + \boldsymbol{\tau}_2)\, \mathbf{dx}.$$

As a result, we obtain the following estimate:

$$\frac{\nu}{2} \|\varepsilon(\mathbf{v} - \mathbf{u})\|^2 \leq \mathcal{M}_1(\mathbf{v}, \boldsymbol{\tau}_{1\mathbf{f}}, \boldsymbol{\tau}_2) := \mathbf{D}_1(\varepsilon(\mathbf{v}), \boldsymbol{\tau}_{1\mathbf{f}}) + \mathbf{D}_2(\varepsilon(\mathbf{v}), \boldsymbol{\tau}_2), \qquad (305)$$

Here we have introduced *compound functionals*

$$\mathbf{D_1}(\varepsilon(\mathbf{v}), \boldsymbol{\tau}_{\mathbf{1f}}) := \int_\Omega \left(\tfrac{\nu}{2} \mid \varepsilon(\mathbf{v}) \mid^2 + \tfrac{1}{2\nu} \mid \boldsymbol{\tau}_{\mathbf{1f}} \mid^2 - \varepsilon(\mathbf{v}) : \boldsymbol{\tau}_{\mathbf{1f}} \right) \mathbf{dx} =$$

$$= \tfrac{1}{2\nu} \| \nu\varepsilon(\mathbf{v}) - \boldsymbol{\tau}_{\mathbf{1f}} \|^2;$$

$$\mathbf{D_2}(\varepsilon(\mathbf{v}), \boldsymbol{\tau}_2) := \int_\Omega (\psi(\varepsilon(\mathbf{v})) + \psi^*(\boldsymbol{\tau}_2) - \varepsilon(\mathbf{v}) : \boldsymbol{\tau}_2) \, \mathbf{dx}.$$

It is clear that both functionals $\mathbf{D_1}$ and $\mathbf{D_2}$ are nonnegative. Moreover,

$$\mathbf{D_1}(\varepsilon(\mathbf{v}), \boldsymbol{\tau}_{\mathbf{1f}}) = \mathbf{0}$$

if and only if

$$\boldsymbol{\tau}_{\mathbf{1f}} = \nu\varepsilon(\mathbf{v}).$$

By the properties of the conjugate functionals

$$\mathbf{D}_2(\varepsilon(\mathbf{v}), \tau_2) = \mathbf{0}$$

if and only if

$$\tau_2 \in \partial\psi(\varepsilon(\mathbf{v})).$$

Now, it is easy to understand the meaning of the estimates (305). Present the main system in the form

$$-\mathbf{div}(\sigma_1 + \sigma_2) = \mathbf{f} - \nabla \mathbf{p} \quad \text{in } \Omega, \tag{306}$$

$$\mathbf{div u} = \mathbf{0} \quad \text{in } \Omega, \tag{307}$$

$$\sigma_1 = \nu\varepsilon(\mathbf{u}), \sigma_2 \in \partial\psi(\varepsilon(\mathbf{u})) \quad \text{in } \Omega, \tag{308}$$

$$\mathbf{u} = \mathbf{u_0} \quad \text{on } \partial\Omega. \tag{309}$$

If $\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}$, and $\boldsymbol{\tau}_{1f} \in \bar{\boldsymbol{\Sigma}}_{\mathbf{f}}(\Omega))$, then

$$-\mathbf{div}(\boldsymbol{\tau}_{1\mathbf{f}} + \boldsymbol{\tau}_2) = \mathbf{f} - \nabla \mathbf{q} \text{ and } \mathbf{div}\mathbf{v} = \mathbf{0},$$

so that for $\mathbf{v}$, $\boldsymbol{\tau}_{1f}$, $\boldsymbol{\tau}_2$ and q the relations (306), (307) and (309) are satisfied.
Our estimate shows that in this case, the energy norm of the deviation from the exact solution is controlled by the quantities

$$\mathbf{D_1}(\varepsilon(\mathbf{v}), \boldsymbol{\tau}_{1\mathbf{f}}) \text{ and } \mathbf{D_2}(\varepsilon(\mathbf{v}), \boldsymbol{\tau}_2),$$

Certainly, the condition $\boldsymbol{\tau}_{1f} \in \bar{\boldsymbol{\Sigma}}_{\mathbf{f}}(\Omega)$ is rather obligatory and it would be useful to somehow eliminate it. This can be done by the same method as we have discussed for linear problems.
As a result, we obtain the estimate

$$\frac{\nu}{2}\|\varepsilon(\mathbf{v}-\mathbf{u})\|^2 \le (1+\beta)\mathbf{D}_1(\varepsilon(v),\tau_1)+\mathbf{D}_2(\varepsilon(v),\tau_2) + \frac{1+\beta}{2\nu\beta}\left[\mathbf{div}(\tau_1+\tau_2)+\mathbf{f}-\nabla\mathbf{q}\right]^2, \qquad (310)$$

that holds for any function $\mathbf{v} \in \mathring{\mathbf{J}}^1_2(\Omega)+\mathbf{u_0}$, any pair of functions $(\boldsymbol{\tau}_1,\boldsymbol{\tau}_2) \in \boldsymbol{\Sigma}\times\boldsymbol{\Sigma}$, and any $\beta>0$. The right–hand side of (310) is the majorant of the norm of the deviation from the exact solution that we denote by $\mathcal{M}_2(\beta,\mathbf{v},\boldsymbol{\tau}_1,\boldsymbol{\tau}_2,\mathbf{q})$.

If the sum $\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2$ has a higher regularity, so that

$$(\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2) \in \boldsymbol{\Sigma}_{\mathbf{div}}(\Omega)$$

and, in addition, $\mathbf{q} \in \mathbf{H}^1$, then the last term of the Majorant is estimated by an explicitly computable integral:

$$\frac{\nu}{2}\|\varepsilon(\mathbf{v}-\mathbf{u})\|^2 \leq (1+\beta)\mathbf{D}_1(\varepsilon(\mathbf{v}),\boldsymbol{\tau}_1) + \mathbf{D}_2(\varepsilon(\mathbf{v}),\boldsymbol{\tau}_2) + \\ + \frac{1+\beta}{2\nu\beta}\mathbf{C}_\Omega^2\|\mathbf{div}(\boldsymbol{\tau}_1+\boldsymbol{\tau}_2) + \mathbf{f} - \nabla\mathbf{q}\|^2. \quad (311)$$

Theorem

For any $\beta \in \mathbb{R}_+$, $\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega) + \mathbf{u_0}$, $\tau_1 \in \mathbf{\Sigma}$, $\tau_2 \in \mathbf{\Sigma}$ and $\mathbf{q} \in \overset{\circ}{\mathbf{L}}{}_2(\Omega)$ the functional $\mathcal{M}_2(\beta, \mathbf{v}, \tau_1, \tau_2, \mathbf{q})$ majorizes the quantity $\|\varepsilon(\mathbf{v} - \mathbf{u})\|^2$.
For any $\beta \in \mathbb{R}_+$, infimum of this functional on the set

$$(\overset{\circ}{\mathbf{J}}{}^1_2(\Omega) + \mathbf{u_0}) \times \mathbf{\Sigma} \times \mathbf{\Sigma} \times \overset{\circ}{\mathbf{L}}{}_2(\Omega)$$

is equal to zero and it is attained if and only if $\mathbf{v} = \mathbf{u}$, $\tau_1 = \sigma_1$, $\tau_2 = \sigma_2$ and $\mathbf{q} = \mathbf{p}$.

Example 1.
For the Stokes problem $\mathbf{K}_* = \mathbf{0}$, $\psi(\boldsymbol{\varepsilon}) \equiv \mathbf{0}$. Set $\boldsymbol{\tau}_2 = 0$. Then, $\mathbf{D}_2(\boldsymbol{\varepsilon}(v), \boldsymbol{\tau}_2) \equiv 0$ and (310) comes in the form we have already obtained.

Example 2.
For the Bingham model $\psi(\boldsymbol{\varepsilon}) = \mathbf{K}_* \mid \boldsymbol{\varepsilon} \mid$, and

$$\psi^*(\boldsymbol{\tau}(\mathbf{x})) = \begin{cases} \mathbf{0}, & \text{if} |\boldsymbol{\tau}(\mathbf{x})| \leq \mathbf{K}_*, \\ +\infty, & \text{if} |\boldsymbol{\tau}(\mathbf{x})| > \mathbf{K}_*. \end{cases}$$

Therefore,

$$\mathbf{D_2}(\varepsilon(\mathbf{v}), \tau_2) = \int_\Omega (\mathbf{k}_* \mid \varepsilon(\mathbf{v}) \mid - \varepsilon(\mathbf{v}) \cdot \tau_2) \mathbf{dx},$$

if almost everywhere the function τ_2 satisfies the condition $\mid \tau_2(\mathbf{x}) \mid \leq \mathbf{K}_*$. In the opposite case, $\mathbf{D_2}(\varepsilon(\mathbf{v}), \tau_2) = +\infty$. Then, (310) comes in the form:

$$\frac{\nu}{2} \|\varepsilon(\mathbf{v} - \mathbf{u})\|^2 \leq \int_\Omega \Big(\frac{(1+\beta)}{2\nu} \mid \nu\varepsilon(\mathbf{v}) - \tau_1 \mid^2 + \mathbf{K}_* \mid \varepsilon(\mathbf{v}) \mid - \varepsilon(\mathbf{v}) : \tau_2 \Big) \mathbf{dx}$$
$$+ \frac{1+\beta}{2\beta\nu} \left[\mathbf{div}(\tau_1 + \tau_2) + \mathbf{f} - \nabla \mathbf{q} \right]^2. \qquad (312)$$

If the functions τ_1, τ_2 and q are taken such that $\mathbf{q} \in \mathbf{H^1}$ $\mathbf{div}(\tau_1 + \tau_2) \in \mathbf{L_2(\Omega)}$, then the last term in (312) is estimated by the integral.

Bingham fluid may have too zones: the *congestion zone* Ω_0 (where $\boldsymbol{\varepsilon}(\mathbf{u}) \equiv \mathbf{0}$) and the *flow zone* Ω_1 (where $\mid \boldsymbol{\varepsilon}(\mathbf{u}) \mid > \mathbf{0}$).
Assume that the right–hand side of (312) vanishes for certain functions $\mathbf{v}$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$ and $\mathbf{q}$. Then, in Ω_0 we have $\boldsymbol{\varepsilon}(\mathbf{v}) = \mathbf{0}$, and, consequently, $\boldsymbol{\tau}_1 = 0$ and $\mathbf{div}\boldsymbol{\tau}_2 + \mathbf{f} - \nabla\mathbf{q} = \mathbf{0}$ for some $\boldsymbol{\tau}_2$, satisfying the condition $\mid \boldsymbol{\tau}_2(\mathbf{x}) \mid \leq \mathbf{1}$.
At the same time, in the flow zone Ω_1 the relations

$$\boldsymbol{\tau}_2 = \mathbf{k}_* \frac{\boldsymbol{\varepsilon}(\mathbf{v})}{\mid \boldsymbol{\varepsilon}(\mathbf{v}) \mid}, \quad \boldsymbol{\tau}_1 = \nu\boldsymbol{\varepsilon}(\mathbf{v}), \quad \mathbf{div}(\boldsymbol{\tau}_1 + \boldsymbol{\tau}_2) + \mathbf{f} - \nabla\mathbf{q} = \mathbf{0}$$

hold.

Lecture 9

GENERAL APPROACH TO A POSTERIORI ERROR CONTROL FOR NONLINEAR VARIATIONAL PROBLEMS

Lecture goal

In subsequent lectures we will present the general theory of a posteriori error control for convex variational problems. In the framework of this theory we are able to derive computable upper bounds for the errors for problems of the type

$$\inf_{\mathbf{v}\in\mathbf{V}} \mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}), \quad \mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}) := \mathbf{G}(\mathbf{\Lambda v}) + \mathbf{F}(\mathbf{v}),$$

where $\mathbf{\Lambda} : \mathbf{V} \longrightarrow \mathbf{Y}$ is a linear continuous operator from a Banach space $\mathbf{V}$ to another Banach space $\mathbf{Y}$ and $\mathbf{J} : \mathbf{Y} \longrightarrow \mathbb{R}$ and $\mathbf{F} : \mathbf{V} \longrightarrow \mathbb{R}$ are convex l.s.c. functionals.

In particular, if

$$\mathbf{\Lambda v} = \nabla \mathbf{v}, \quad \mathbf{G(y)} = (\mathbf{Ay}, \mathbf{y}), \quad \mathbf{F(v)} = (\mathbf{f}, \mathbf{v}),$$

then we arrive to the variational formulation of the problem

$$\mathbf{div\, A} \nabla \mathbf{u} + \mathbf{f} = \mathbf{0}.$$

Many other problems have the above form, were

G is the energy functional whose form is dictated by the dissipative properties of a media.
F is the functional associated with external forces and (or) boundary conditions.

Diffusion type problems,
Linear elasticity,
Biharmonic problems,
Kirhghoff and Mindlin plates,
Problems in deformation theory of elastoplasticity,
p-Laplace equation,
Stokes problem,
Nonlinear problems in the theory of viscous fluids and many other problems can be presented in the above general form.

In such models, the structure of the "energy functional" G plays crucial role in all the parts of the mathematical analysis: existence and differentiability properties of minimizers and estimates of deviations from the minimizers.

To understand the basic principles of the functional approach to the derivation of a posteriori bounds of the approximation errors we need to make a concise overview of some parts of the duality theory in the calculus of variations.

A consequent exposition functional type a posteriori error estimates for nonlinear variational problems can be found in the papers

S. Repin. A posteriori error estimation for variational problems with uniformly convex functionals, *Math. Comput.*, 69(230), 2000, 481-500.

S. Repin. Two-sided estimates for deviation from an exact solution to uniformly elliptic equation. *Trudi St.-Petersburg Math. Society*, 9(2001), 148-179 (in Russian, translated in *American Mathematical Translations Series 2, 9(2003)*

and in the book

P. Neittaanmaki and S. Repin. Reliable methods for computer simulation. Error control and a posteriori estimates. Elsevier, NY, 2004.

Selected topics of the duality theory in the calculus of variations

To understand the structure of functional a posteriori estimates for the considered class of problems we need first discuss three additional topics:

- **Dual and bidual functionals ;**
- **Compound functionals ;**
- **Uniformly convex functionals.**

Dual (polar) functionals

Hereafter $\mathbf{V}^*$ contains all linear continuous functionals defined on $\mathbf{V}$. The elements of $\mathbf{V}^*$ are marked by stars,
$\langle \mathbf{v}^*, \mathbf{v} \rangle$ is called the **duality pairing** of the spaces $\mathbf{V}$ and $\mathbf{V}^*$.
Let $\mathbf{J} : \mathbf{V} \longrightarrow \mathbb{R}$, then $\mathbf{J}^*$ defined by the relation

$$\mathbf{J}^*(\mathbf{v}^*) = \sup_{\mathbf{v} \in \mathbf{V}} \{ \langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}(\mathbf{v}) \}$$

is called **dual** to $\mathbf{J}$.
If $\mathbf{J}$ is a smooth function that increases at infinity faster than any linear function, then $\mathbf{J}^*$ is the Legendre transform of $\mathbf{J}$. The above general definition comes from Young and Fenchel. The functional $\mathbf{J}^*$ is also called **polar** to $\mathbf{J}$.

Bipolar functionals

The functional

$$\mathbf{J}^{**}(\mathbf{v}) = \sup_{\mathbf{v}^* \in \mathbf{V}^*} \{\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}^*(\mathbf{v}^*)\}$$

is called the **bidual** to $\mathbf{J}$ (or **bipolar**).
Straightforwardly from the definition, it follows that $\mathbf{J}^*$ and $\mathbf{J}^{**}$ are convex functionals (they are defined as upper bounds of affine functionals).
Formally, one can also define

$$\mathbf{J}^{***}(\mathbf{v}^*) := \sup_{\mathbf{v} \in \mathbf{V}} \{\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}^{**}(\mathbf{v})\}.$$

However, this definition brings nothing new. It is proved that

$$\mathbf{J}^{***}(\mathbf{v}^*) = \mathbf{J}^*(\mathbf{v}^*), \quad \forall \mathbf{v}^* \in \mathbf{V}^*.$$

Mutually dual functionals

Let $\mathbf{J} : \mathbf{V} \to \overline{\mathbb{R}} := \{\mathbb{R}, -\infty, +\infty\}$ and $\mathbf{G}^* : \mathbf{V}^* \to \overline{\mathbb{R}}$ be two functionals defined on a Banach space $\mathbf{V}$ and its dual space $\mathbf{V}^*$, respectively. These two functionals are called **mutually dual** if

$$(\mathbf{G}^*)^* = \mathbf{J} \quad \text{and} \quad \mathbf{J}^* = \mathbf{G}^*.$$

Examples

To illustrate the definitions of conjugate functionals, we present below several examples for functionals defined on the Euclidean space $\mathbf{E^d}$. In this case, $\mathbf{V}$ and $\mathbf{V}^*$ are isometrically isomorphic. Their elements are d-dimensional vectors denoted by $\boldsymbol{\xi}$ and $\boldsymbol{\xi}^*$, respectively, so that

$$\langle \boldsymbol{\xi}^*, \boldsymbol{\xi} \rangle = \boldsymbol{\xi}^* \cdot \boldsymbol{\xi} = \boldsymbol{\xi}^*_i \boldsymbol{\xi}_i .$$

These examples have a practical meaning because for a wide class of integral type functionals (in the mechanics they are the **energy functionals**) finding the dual energy functional is reduced to **finding dual to its integrand !**

In other words, if the "primal energy functional" has the form

$$\mathbf{G(v)} := \int_{\Omega} \mathbf{g(\Lambda v)dx}$$

where $\mathbf{g}$ is the "internal energy" or "dissipative potential", then the so–called "complementary energy" is given by the integral functional

$$\mathbf{G^*(y^*)} := \int_{\Omega} \mathbf{g^*(y^*)dx},$$

where $\mathbf{g}^*$ is conjugate to $\mathbf{g}$ in the algebraic sense.

Example 1 (Diffusion problems)

Let $\mathbf{A} = \{\mathbf{a_{ij}}\}$ be a real, positive definite matrix and

$$\mathbf{g}(\xi) = \frac{1}{2}\mathbf{A}\xi \cdot \xi = \frac{1}{2}\mathbf{a_{ij}}\xi_i\xi_j.$$

Then

$$\mathbf{g}^*(\xi^*) = \sup_{\xi \in \mathbf{E^d}} \left\{ \xi^* \cdot \xi - \frac{1}{2}\mathbf{A}\xi \cdot \xi \right\}.$$

This supremum is attained on an element ξ_0 such that

$$\xi^* = \mathbf{A}\xi_0 \implies \xi_0 = \mathbf{A}^{-1}\xi^*.$$

Therefore, we have a pair of mutually conjugate functionals

$$\mathbf{g}(\xi) = \frac{1}{2}\mathbf{A}\xi \cdot \xi \quad \text{and} \quad \mathbf{g}^*(\xi^*) = \frac{1}{2}\mathbf{A}^{-1}\xi^* \cdot \xi^*.$$

In diffusion type boundary–value problems we arrive at the functional (with $\mathbf{y} = \nabla \mathbf{v}$)

$$\frac{1}{2} \int_{\Omega} \mathbf{A}\mathbf{y} \cdot \mathbf{y} \, \mathbf{dx} \quad \mathbf{y} \in \mathbf{L}^2(\Omega, \mathbb{R}^n),$$

which is mutually dual to

$$\frac{1}{2} \int_{\Omega} \mathbf{A}^{-1}\mathbf{y}^* \cdot \mathbf{y}^* \, \mathbf{dx} \quad \mathbf{y}^* \in \mathbf{L}^2(\Omega, \mathbb{R}^n)$$

Example 2 (Linear elasticity)

Let $\mathbf{L} = \{\mathbf{L_{ijkm}}\}$ be a real, positive definite tensor of the 4-th order and $\boldsymbol{\tau}$ be a tensor of the second order ($\mathbf{d} \times \mathbf{d}$–matrix). Then,

$$\mathbf{g}(\varepsilon) = \frac{1}{2}\mathbf{L}\varepsilon : \varepsilon = \frac{1}{2}\mathbf{L_{ijkm}}\varepsilon_{\mathbf{ij}}\varepsilon_{\mathbf{km}}.$$

Then

$$\mathbf{g}^*(\varepsilon^*) = \sup_{\varepsilon \in \mathbf{M^{d\times d}}} \left\{ \varepsilon^* : \varepsilon - \frac{1}{2}\mathbf{A}\varepsilon : \varepsilon \right\}.$$

This supremum is attained on an element ε_0 such that

$$\boldsymbol{\tau}^* = \mathbf{L}\varepsilon_0 \implies \varepsilon_0 = \mathbf{L}^{-1}\varepsilon^*.$$

Therefore, we have a pair of mutually dual functionals

$$\mathbf{g}(\varepsilon) = \frac{1}{2}\mathbf{L}\varepsilon : \varepsilon \quad \text{and} \quad \mathbf{g}^*(\varepsilon^*) = \frac{1}{2}\mathbf{L}^{-1}\varepsilon^* : \varepsilon^*.$$

In linear elasticity problems we arrive at the energy functional in terms of strains $\varepsilon(\mathbf{v}) = \frac{1}{2}(\nabla \mathbf{v} + (\nabla \mathbf{v})^{\mathsf{T}})$

$$\frac{1}{2}\int_{\Omega} \mathbb{L}\varepsilon : \varepsilon \, \mathbf{dx} \quad \varepsilon \in \mathbf{L}^2(\Omega, \mathbb{M}^{n\times n}),$$

which is mutually dual to the "complementary energy" functional written in terms of stresses $\varepsilon^*(x) => \tau(x)$

$$\frac{1}{2}\int_{\Omega} \mathbb{L}^{-1}\tau : \tau \, \mathbf{dx} \quad \tau \in \mathbf{L}^2(\Omega, \mathbb{M}^{n\times n})$$

Example 3 (Nonlinear elasticity, p-Laplacian)

Consider the functional

$$\mathbf{g}(\boldsymbol{\xi}) = \frac{\mathbf{1}}{\mathbf{p}}|\boldsymbol{\xi}|^{\mathbf{p}},$$

where $\mathbf{p} > \mathbf{1}$ and $|\boldsymbol{\xi}| = (\boldsymbol{\xi} \cdot \boldsymbol{\xi})^{1/2}$. It is easy to verify that the quantity $\boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - \frac{\mathbf{1}}{\mathbf{p}}|\boldsymbol{\xi}|^{\mathbf{p}}$ attains a supremum if $\boldsymbol{\xi} = \boldsymbol{\xi}_0$, where $\boldsymbol{\xi}_0$ satisfies the relation

$$\boldsymbol{\xi}^* - |\boldsymbol{\xi}_0|^{\mathbf{p}-\mathbf{2}}\boldsymbol{\xi}_0 = \mathbf{0},$$

which yields $|\boldsymbol{\xi}^*| = |\boldsymbol{\xi}_0|^{\mathbf{p}-\mathbf{1}}$ and $\boldsymbol{\xi}^* \cdot \boldsymbol{\xi}_0 = |\boldsymbol{\xi}_0|^{\mathbf{p}}$. Therefore,

$$\mathbf{g}^*(\boldsymbol{\xi}^*) = \boldsymbol{\xi}^* \cdot \boldsymbol{\xi}_0 - \frac{\mathbf{1}}{\mathbf{p}}|\boldsymbol{\xi}_0|^{\mathbf{p}} = \left(\mathbf{1} - \frac{\mathbf{1}}{\mathbf{p}}\right)|\boldsymbol{\xi}_0|^{\mathbf{p}} = \frac{\mathbf{1}}{\mathbf{p}^*}|\boldsymbol{\xi}^*|^{\mathbf{p}^*},$$

where $\mathbf{p}^* = \frac{\mathbf{p}}{\mathbf{p}-\mathbf{1}}$.

Thus, we obtain another pair of mutually conjugate functionals

$$\mathbf{g}(\xi) = \frac{1}{\mathbf{p}}|\xi|^{\mathbf{p}} \quad \text{and} \quad \mathbf{g}^*(\xi^*) = \frac{1}{\mathbf{p}^*}|\xi^*|^{\mathbf{p}^*},$$

where $\frac{1}{\mathbf{p}} + \frac{1}{\mathbf{p}^*} = 1$.

Remark

This relation admits generalizations. Namely, let $\varphi : \mathbb{R} \to \mathbb{R}$ be a proper convex function that is, in addition, odd and let $\varphi^* : \mathbb{R} \to \mathbb{R}$ be its conjugate. Then

$$(\varphi(\|\mathbf{u}\|\mathbf{v}))^* = \varphi^*(\|\mathbf{u}^*\|\mathbf{v}^*).$$

In certain nonlinear boundary–value problems we arrive at the functional (with $\mathbf{y} = \nabla \mathbf{v}$ or $\mathbf{y} = \boldsymbol{\varepsilon}(\mathbf{v})$)

$$\frac{1}{p} \int_{\Omega} |\mathbf{y}|^{p} \, dx \quad \mathbf{y} \in \mathbf{L}^{p}(\Omega, \mathbb{R}^{n}[\mathbb{M}^{n \times n}]),$$

which is mutually dual to

$$\frac{1}{p^{*}} \int_{\Omega} |\mathbf{y}^{*}|^{p^{*}} \, dx \quad \mathbf{y}^{*} \in \mathbf{L}^{p^{*}}(\Omega, \mathbb{R}^{n}[\mathbb{M}^{n \times n}]).$$

Example 4 (Action of external forces)

Let $\mathbf{g}(\boldsymbol{\xi})$ be a linear functional, i.e.,

$$\mathbf{g}(\boldsymbol{\xi}) = \ell \cdot \boldsymbol{\xi}, \quad \ell \in \mathbf{E^d}.$$

It is easy to see that

$$\mathbf{g}^*(\boldsymbol{\xi}^*) = \sup_{\boldsymbol{\xi} \in \mathbf{E^d}} \{\boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - \ell \cdot \boldsymbol{\xi}\} = \begin{cases} 0 & \boldsymbol{\xi}^* = \ell, \\ +\infty & \boldsymbol{\xi}^* \neq \ell. \end{cases}$$

Denote by $\mathfrak{X}_{\{\ell\}}$ the characteristic functional of the set $\{\ell\} \subset \mathbf{E^d}$. Then, another pair of mutually conjugate functionals is as follows:

$$\mathbf{g}(\boldsymbol{\xi}) = \ell \cdot \boldsymbol{\xi} \quad \text{and} \quad \mathbf{g}^*(\boldsymbol{\xi}^*) = \mathfrak{X}_{\{\ell\}}(\boldsymbol{\xi}^*).$$

Thus, for the functional $\mathbf{G} : \mathbf{L}^2 \longrightarrow \mathbb{R}$

$$\mathbf{G}(\mathbf{v}) := \int_{\Omega} \mathbf{f}\mathbf{v}\,\mathbf{dx},\ \mathbf{f} \in \mathbf{L}^2(\Omega)$$

the respective dual functional is $\mathbf{G}^* : \mathbf{L}^2 \longrightarrow \mathbb{R}$

$$\mathbf{G}^*(\mathbf{v}^*) = \mathbf{0} \text{ if } \mathbf{v}^* = \mathbf{f}, \qquad \mathbf{G}^*(\mathbf{v}^*) = +\infty \text{ in other cases.}$$

Example 5 (Variational inequalities)

Let $\mathbf{g}(\boldsymbol{\xi}) = |\boldsymbol{\xi}|$. Then

$$\sup_{\boldsymbol{\xi}} \{\boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - |\boldsymbol{\xi}|\}$$

may be finite or infinite depending on the value of $|\boldsymbol{\xi}^*|$. If $|\boldsymbol{\xi}^*| > 1$, then, obviously, it is infinite. If $|\boldsymbol{\xi}^*| \leq 1$, then, on the one hand,

$$\sup_{\boldsymbol{\xi}} \{\boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - |\boldsymbol{\xi}|\} \leq \sup_{\boldsymbol{\xi}} \{\mathbf{1}|\boldsymbol{\xi}| - |\boldsymbol{\xi}|\} = \mathbf{0}.$$

On the other hand, $\sup_{\boldsymbol{\xi}} \{\boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - |\boldsymbol{\xi}|\} \geq \boldsymbol{\xi}^* \cdot \mathbf{0} - \mathbf{0} = \mathbf{0}$. This means that $\mathbf{g}^*(\boldsymbol{\xi}^*) = \mathbf{0}$ if $|\boldsymbol{\xi}^*| \leq 1$ and, thus,

$$\mathbf{g}(\boldsymbol{\xi}) = |\boldsymbol{\xi}|, \ \mathbf{g}^*(\boldsymbol{\xi}^*) = \mathfrak{X}_{\mathcal{B}^*(\mathbf{0},1)}(\boldsymbol{\xi}^*), \text{ where } \mathcal{B}^*(\mathbf{0},1) = \{\boldsymbol{\xi}^* \in \mathbf{E}^{\mathbf{d}} \mid |\boldsymbol{\xi}^*| \leq \mathbf{1}\}.$$

Thus, for the functional $\mathbf{G} : \mathbf{L}^1 \to \mathbb{R}$

$$\mathbf{G}(\mathbf{v}) := \int_{\Omega} |\mathbf{v}| \, \mathbf{dx},$$

the respective dual functional is $\mathbf{G}^* : \mathbf{L}^\infty \to \mathbb{R}$

$$\mathbf{G}^*(\mathbf{v}^*) = \mathbf{0} \text{ if } |\mathbf{v}^*(\mathbf{x})| \leq \mathbf{1} \text{ a.e. in } \Omega, \quad \mathbf{G}^*(\mathbf{v}^*) = +\infty \text{ in other cases.}$$

Properties of dual functionals

Property 1

If $\mathbf{J} : \mathbf{V} \to \overline{\mathbb{R}}$ and $\mathbf{G} : \mathbf{V} \to \overline{\mathbb{R}}$ are such that

$$\mathbf{J}(\mathbf{v}) \geq \mathbf{G}(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V},$$

then

$$\mathbf{J}^*(\mathbf{v}^*) \leq \mathbf{G}^*(\mathbf{v}^*), \quad \forall \mathbf{v}^* \in \mathbf{V}^*.$$

Proof. We have

$$\mathbf{J}^*(\mathbf{v}^*) = \sup_{\mathbf{v} \in \mathbf{V}} \{\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}(\mathbf{v})\} \leq \sup_{\mathbf{v} \in \mathbf{V}} \{\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{G}(\mathbf{v})\} = \mathbf{G}^*(\mathbf{v}^*).$$

Property 2

For any $\lambda > 0$,

$$(\lambda \mathbf{J})^*(\mathbf{v}^*) = \lambda \mathbf{J}^* \left(\frac{\mathbf{v}^*}{\lambda} \right).$$

Proof. This property is justified by direct calculations:

$$(\lambda \mathbf{J})^*(\mathbf{v}^*) = \sup_{\mathbf{v} \in \mathbf{V}} \{ \langle \mathbf{v}^*, \mathbf{v} \rangle - \lambda \mathbf{J}(\mathbf{v}) \} =$$

$$= \lambda \sup_{\mathbf{v} \in \mathbf{V}} \left\{ \left\langle \frac{\mathbf{v}^*}{\lambda}, \mathbf{v} \right\rangle - \mathbf{J}(\mathbf{v}) \right\} = \lambda \mathbf{J}^* \left(\frac{\mathbf{v}^*}{\lambda} \right).$$

Property 3

Let $\mathbf{J} : \mathbf{V} \to \overline{\mathbb{R}}$ and $\mathbf{J}_\alpha(\mathbf{v}) = \mathbf{J}(\mathbf{v}) + \alpha$, where $\alpha \in \mathbb{R}$. Then

$$\mathbf{J}^*_\alpha(\mathbf{v}^*) = \mathbf{J}^*(\mathbf{v}^*) - \alpha.$$

Proof. It follows from the obvious relation

$$\sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J}(\mathbf{v}) - \alpha\} = \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J}(\mathbf{v})\} - \alpha.$$

Property 4

Let $\mathbf{v_0} \in \mathbf{V}$ and $\mathbf{G(v)} = \mathbf{J(v - v_0)}$. Then

$$\mathbf{G^*(v^*)} = \mathbf{J^*(v^*)} + \langle \mathbf{v^*}, \mathbf{v_0} \rangle.$$

Proof. Since

$$\sup_{\mathbf{v} \in \mathbf{V}} \{\langle \mathbf{v^*}, \mathbf{v} \rangle - \mathbf{J(v - v_0)}\} = \sup_{\mathbf{w} \in \mathbf{V}} \{\langle \mathbf{v^*}, \mathbf{w + v_0} \rangle - \mathbf{J(w)}\}$$
$$= \sup_{\mathbf{w} \in \mathbf{V}} \{\langle \mathbf{v^*}, \mathbf{w} \rangle - \mathbf{J(w)}\} + \langle \mathbf{v^*}, \mathbf{v_0} \rangle = \mathbf{J^*(v^*)} + \langle \mathbf{v^*}, \mathbf{v_0} \rangle,$$

we arrive at the required relation.

Property 5

If $\mathbf{G}(\mathbf{v}) = \min_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}$, then $\mathbf{G}^*(\mathbf{v}^*) = \max_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i^*}(\mathbf{v}^*)\}$.

Proof. We have

$$\begin{aligned}
\mathbf{G}^*(\mathbf{v}^*) &= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \min_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}\} \\
&= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle + \max_{\mathbf{i}=1,\dots,\mathbf{N}}\{-\mathbf{J_i}(\mathbf{v})\}\} \\
&= \sup_{\mathbf{v}\in\mathbf{V}} \max_{\mathbf{i}=1,\dots,\mathbf{N}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\} \\
&= \max_{\mathbf{i}=1,\dots,\mathbf{N}} \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\} = \max_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i^*}(\mathbf{v}^*)\}.
\end{aligned}$$

Property 6

If $\mathbf{G}(\mathbf{v}) = \max_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}$, then $\mathbf{G}^*(\mathbf{v}^*) \le \min_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i^*}(\mathbf{v}^*)\}$.

Proof. By definition, we have

$$\begin{aligned}\mathbf{G}^*(\mathbf{v}^*) &= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \max_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}\}\\ &= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle + \min_{\mathbf{i}=1,\dots,\mathbf{N}}\{-\mathbf{J_i}(\mathbf{v})\}\}\\ &= \sup_{\mathbf{v}\in\mathbf{V}}\min_{\mathbf{i}=1,\dots,\mathbf{N}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\}.\end{aligned}$$

Now we apply $\sup\inf \le \inf\sup$ relation to $\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})$. Then,

$$\mathbf{G}^*(\mathbf{v}^*) \le \min_{\mathbf{i}=1,\dots,\mathbf{N}}\sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\} = \min_{\mathbf{i}=1,\dots,\mathbf{N}}\{\mathbf{J_i^*}(\mathbf{v}^*)\}.$$

Subdifferential

Definition

The functional $\mathbf{J}\mathbf{V} \to \mathbb{R}$ is called subdifferentiable at $\mathbf{v_0}$ if there exists an affine minorant $\ell \in \mathbb{AM}(\mathbf{J})$ such that $\mathbf{J}(\mathbf{v_0}) = \ell(\mathbf{v_0})$. A minorant with this property is called the **exact minorant** at $\mathbf{v_0}$.

Obviously, any affine minorant exact for $\mathbf{J}$ at v_0 has the form

$$\ell(\mathbf{v}) = \langle \mathbf{v}^*, \mathbf{v} - \mathbf{v_0} \rangle + \mathbf{J}(\mathbf{v_0}), \quad \ell(\mathbf{v}) \leq \mathbf{J}(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V}.$$

The element $\mathbf{v}^*$ is called a **subgradient** of $\mathbf{J}$ at $\mathbf{v_0}$.

The set of all subgradients of $\mathbf{J}$ at $\mathbf{v_0}$ forms a **subdifferential**, which is usually denoted by $\partial \mathbf{J}(\mathbf{v_0})$. It may be empty or contain one element or infinitely many elements.
An important property of convex functionals follows directly from the above definition. For a convex functional $\mathbf{J}$ at a point $\mathbf{v_0}$ where it is finite, the exact affine minorant is evidently exist!
In other words, there is at least one element $\mathbf{v}^* \in \partial \mathbf{J}(\mathbf{v_0})$ that "creates" an affine minorant such that

$$\langle \mathbf{v}^*, \mathbf{v} \rangle - \boldsymbol{\alpha} \leq \mathbf{J}(\mathbf{v}), \qquad \forall \mathbf{v} \in \mathbf{V},$$
$$\langle \mathbf{v}^*, \mathbf{v_0} \rangle - \boldsymbol{\alpha} = \mathbf{J}(\mathbf{v_0}).$$

By subtracting, we obtain

$$\mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{v_0}) \geq \langle \mathbf{v}^*, \mathbf{v} - \mathbf{v_0} \rangle.$$

The inequality (313) presents the **basic incremental relation for convex functionals**.

Compound functionals

Let $\mathbf{J}$ and $\mathbf{J}^*$ be a pair of mutually dual convex functionals.

The functional $\mathbf{D_J} : \mathbf{V} \times \mathbf{V}^* \longrightarrow \mathbb{R}$ of the form

$$\mathbf{D_J}(\mathbf{v}, \mathbf{v}^*) := \mathbf{J}(\mathbf{v}) + \mathbf{J}^*(\mathbf{v}^*) - \langle \mathbf{v}^*, \mathbf{v} \rangle.$$

is called it the **compound functional** associated with these pair of functionals.

We will see that compound functionals play an important role in the a posteriori analysis of linear and nonlinear variational problems.

Compound functionals are always nonnegative. Indeed,

$$\mathbf{J}^*(\mathbf{v}^*) = \sup_{\mathbf{v}\in\mathbf{V}} \left(\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J}(\mathbf{v})\right) \geq \langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J}(\mathbf{v}) \quad \forall \mathbf{v} \in \mathbf{V}$$

and

$$\mathbf{J}^*(\mathbf{v}^*) + \mathbf{J}(\mathbf{v}) - \langle \mathbf{v}^*, \mathbf{v}\rangle \geq \mathbf{0} \qquad \forall \mathbf{v},\, \mathbf{v}^*$$

Compound functionals may vanish only on special sets, where $\mathbf{v}$ and $\mathbf{v}^*$ satisfy certain relations.

Theorem

Let $\mathbf{J}$ be a proper convex functional and $\mathbf{J}^*$ be its polar. Then, the following two statements are equivalent:

$$\mathbf{J}(\mathbf{v}) + \mathbf{J}^*(\mathbf{v}^*) - \langle \mathbf{v}^*, \mathbf{v} \rangle = \mathbf{0}, \tag{313}$$

$$\mathbf{v}^* \in \partial \mathbf{J}(\mathbf{v}) \quad \text{and} \quad \mathbf{v} \in \partial \mathbf{J}^*(\mathbf{v}^*). \tag{314}$$

Relations (314) are also called **duality relations** for the pair $(\mathbf{v}, \mathbf{v}^*)$.

Proof.

Assume that $\mathbf{v}^* \in \partial \mathbf{J}(\mathbf{v})$., i.e,

$$\mathbf{J}(\mathbf{w}) \geq \mathbf{J}(\mathbf{v}) + \langle \mathbf{v}^*, \mathbf{w} - \mathbf{v} \rangle, \qquad \forall \mathbf{w} \in \mathbf{V}.$$

Hence,

$$\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}(\mathbf{v}) \geq \langle \mathbf{v}^*, \mathbf{w} \rangle - \mathbf{J}(\mathbf{w}), \qquad \forall \mathbf{w} \in \mathbf{V}$$

and, consequently,

$$\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}(\mathbf{v}) \geq \sup_{\mathbf{w} \in \mathbf{V}} \{ \langle \mathbf{v}^*, \mathbf{w} \rangle - \mathbf{J}(\mathbf{w}) \} = \mathbf{J}^*(\mathbf{v}^*),$$

what leads to the conclusion that $\mathbf{J}^*(\mathbf{v}^*) + \mathbf{J}(\mathbf{w}) - \langle \mathbf{v}^*, \mathbf{w} \rangle \leq \mathbf{0}$.
But the left–hand side is nonnegative, so that we obtain
$\mathbf{D}_{\mathbf{J}}(\mathbf{v}^*, \mathbf{v}) = \mathbf{0}$.

Assume that $\mathbf{v} \in \partial \mathbf{J}^*(\mathbf{v}^*)$. Then

$$\mathbf{J}^*(\mathbf{w}^*) \geq \mathbf{J}^*(\mathbf{v}^*) + \langle \mathbf{w}^* - \mathbf{v}^*, \mathbf{v} \rangle,$$

and we continue similarly to the previous case:

$$\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}^*(\mathbf{v}^*) \geq \langle \mathbf{w}^*, \mathbf{v} \rangle - \mathbf{J}^*(\mathbf{w}^*), \quad \forall \mathbf{w}^* \in \mathbf{V}^*,$$
$$\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}^*(\mathbf{v}^*) \geq \mathbf{J}^{**}(\mathbf{v}) = \mathbf{J}(\mathbf{v}).$$

Thus, we again arrive at the conclusion that it can only be if $\mathbf{D}_\mathbf{J}(\mathbf{v}^*, \mathbf{v}) = \mathbf{0}$.

Assume that $\mathbf{D_J}(\mathbf{v}^*, \mathbf{v}) = \mathbf{0}$. Since

$$\mathbf{J}^*(\mathbf{v}^*) = \sup_{\mathbf{w}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{w}\rangle - \mathbf{J}(\mathbf{w})\},$$

we obtain

$$\mathbf{0} = \mathbf{J}(\mathbf{v}) + \mathbf{J}^*(\mathbf{v}^*) - \langle \mathbf{v}^*, \mathbf{v}\rangle \geq \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{w}) - \langle \mathbf{v}^*, \mathbf{v} - \mathbf{w}\rangle, \ \forall \mathbf{w} \in \mathbf{V}.$$

Rewrite this inequality in a more familiar form:

$$\mathbf{J}(\mathbf{w}) - \mathbf{J}(\mathbf{v}) \geq \langle \mathbf{v}^*, \mathbf{w} - \mathbf{v}\rangle, \qquad \forall \mathbf{w} \in \mathbf{V},$$

which means that $\mathbf{J}(\mathbf{v}) + \langle \mathbf{v}^*, \mathbf{v} - \mathbf{w}\rangle$ is an exact affine minorant of $\mathbf{J}$ (at $\mathbf{v}$) and, consequently, $\mathbf{v}^* \in \partial \mathbf{J}(\mathbf{v})$. The proof of the fact that $\mathbf{v}^* \in \partial \mathbf{J}^*(\mathbf{v}^*)$ is quite analogous.

Properties of compound functionals

First, we note that, $\mathbf{D_G(y, y^*)}$ is convex with respect to $\mathbf{y}$ and $\mathbf{y^*}$, but, in general, $\mathbf{D_G(y, y^*)}$ is a nonconvex functional on $\mathbf{Y} \times \mathbf{Y^*}$.
This fact is easily observed in the simplest case $\mathbf{Y} = \mathbb{R}$ if set

$$\mathbf{G(y)} = \frac{\mathbf{1}}{\boldsymbol{\alpha}}|\mathbf{y}|^{\boldsymbol{\alpha}} \qquad \mathbf{G^*(y)} = \frac{\mathbf{1}}{\boldsymbol{\alpha}^*}|\mathbf{y}|^{\boldsymbol{\alpha}^*}.$$

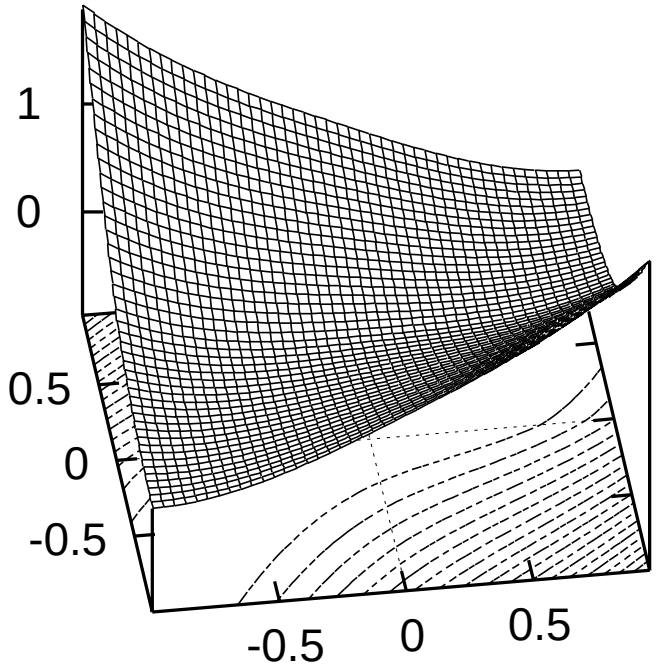
Only for $\boldsymbol{\alpha} = 2$ we have a convex functional

$$\mathbf{D_G(y, y^*)} = \frac{\mathbf{1}}{\mathbf{2}}|\mathbf{y}|^2 + \frac{\mathbf{1}}{\mathbf{2}}|\mathbf{y^*}|^2 - \mathbf{yy^*} = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{y} - \mathbf{y^*})^2.$$

For other $\boldsymbol{\alpha} \in (1, +\infty)$ $\mathbf{D_G}$ is nonconvex on $\mathbb{R} \times \mathbb{R}$.

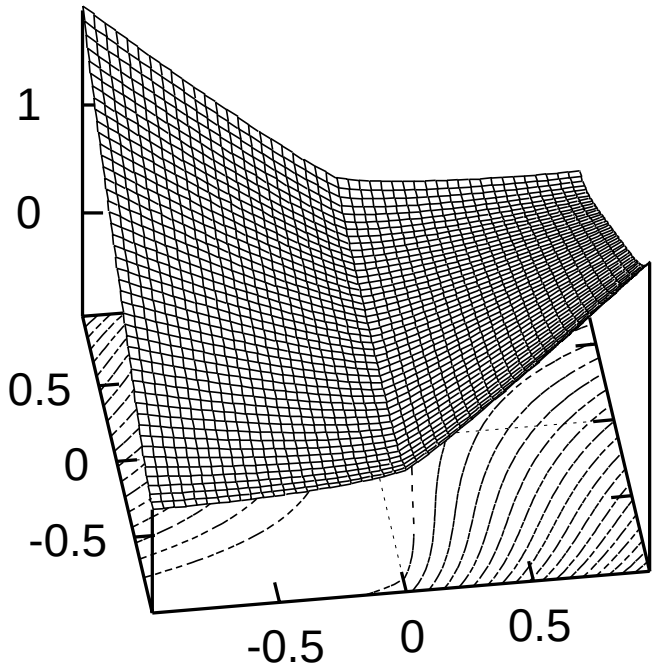
Example 1: $\mathsf{D}(\xi_1, \xi_2) = \frac{1}{3}|\xi_1|^3 + \frac{2}{3}|\xi_2|^{3/2} - \xi_1\xi_2$

Compound functional on $\mathbb{R} \times \mathbb{R}$ and its level lines



Example 2: $\mathsf{D}(\xi_1, \xi_2) = \frac{5}{6}|\xi_1|^{6/5} + \frac{1}{6}|\xi_2|^6 - \xi_1\xi_2$

Compound functional on $\mathbb{R} \times \mathbb{R}$ and its level lines



However, they have an important property, which is to some extent similar to convexity.

Theorem

For any $\mathbf{y}_1, \mathbf{y}_2 \in \mathbf{Y}$ and $\mathbf{y}_1^*, \mathbf{y}_2^* \in \mathbf{Y}^*$,

$$\mathbf{D_G}\left(\tfrac{\mathbf{y}_1+\mathbf{y}_2}{2}, \tfrac{\mathbf{y}_1^*+\mathbf{y}_2^*}{2}\right) \leq \tfrac{1}{4}\Big(\mathbf{D_G}(\mathbf{y}_1, \mathbf{y}_1^*) + \mathbf{D_G}(\mathbf{y}_1, \mathbf{y}_2^*) + \\ + \mathbf{D_G}(\mathbf{y}_2, \mathbf{y}_1^*) + \mathbf{D_G}(\mathbf{y}_2, \mathbf{y}_2^*)\Big)$$

Proof

From the definition it follows that

$$\begin{aligned} \mathsf{D}_{\mathsf{G}}\left(\mathsf{y}, \tfrac{\mathsf{y}_1^*+\mathsf{y}_2^*}{2}\right) = \quad & \mathsf{G}(\mathsf{y}) + \mathsf{G}^*\left(\tfrac{\mathsf{y}_1^*+\mathsf{y}_2^*}{2}\right) - \left\langle \tfrac{\mathsf{y}_1^*+\mathsf{y}_2^*}{2}, \mathsf{y} \right\rangle \\ & \leq \tfrac{1}{2}\left(\mathsf{D}_{\mathsf{G}}(\mathsf{y}, \mathsf{y}_1^*) + \mathsf{D}_{\mathsf{G}}(\mathsf{y}, \mathsf{y}_2^*)\right) \end{aligned}$$

and

$$\begin{aligned} \mathsf{D}_{\mathsf{G}}\left(\tfrac{\mathsf{y}_1+\mathsf{y}_2}{2}, \mathsf{y}^*\right) = \quad & \mathsf{G}\left(\tfrac{\mathsf{y}_1+\mathsf{y}_2}{2}\right) + \mathsf{G}^*(\mathsf{y}^*) - \left\langle \mathsf{y}^*, \tfrac{\mathsf{y}_1+\mathsf{y}_2}{2} \right\rangle \\ & \leq \tfrac{1}{2}\left(\mathsf{D}_{\mathsf{G}}(\mathsf{y}_1, \mathsf{y}^*) + \mathsf{D}_{\mathsf{G}}(\mathsf{y}_2, \mathsf{y}^*)\right). \end{aligned}$$

Therefore,

$$\mathsf{D}_{\mathsf{G}}\left(\tfrac{\mathsf{y}_1+\mathsf{y}_2}{2}, \tfrac{\mathsf{y}_1^*+\mathsf{y}_2^*}{2}\right) \leq \tfrac{1}{2}\left(\mathsf{D}_{\mathsf{G}}(\mathsf{y}_1, \tfrac{\mathsf{y}_1^*+\mathsf{y}_2^*}{2}) + \mathsf{D}_{\mathsf{G}}(\mathsf{y}_2, \tfrac{\mathsf{y}_1^*+\mathsf{y}_2^*}{2})\right).$$

and we arrive at the required estimate.

Important property

If $\mathbf{G}$ and $\mathbf{G}^*$ are Gateaux differentiable, then

$$\langle \mathbf{y}^* - \mathbf{G}'(\mathbf{y}), \mathbf{G}^{*\prime}(\mathbf{y}^*) - \mathbf{y} \rangle \quad \geq \quad \mathbf{D}_{\mathbf{G}}(\mathbf{y}, \mathbf{y}^*).$$

Note, that from this relation we conclude that $\mathbf{D}_{\mathbf{J}}$ vanishes if the duality relations are satisfied.

Uniformly convex functionals

Let a proper l.s.c. functional $\boldsymbol{\Upsilon} : \mathbf{Y} \to \overline{\mathbb{R}}$ be subject to the conditions

$$\boldsymbol{\Upsilon}(\mathbf{y}) \geq \mathbf{0}, \ \forall \mathbf{y} \in \mathbf{Y}, \qquad \boldsymbol{\Upsilon}(\mathbf{y}) = \mathbf{0} \iff \mathbf{y} = \mathbf{0}_{\mathbf{Y}}.$$

Definition

A convex functional $\mathbf{J} : \mathbf{Y} \to \overline{\mathbb{R}}$ is called uniformly convex in $\mathcal{B}(0_{\mathbf{Y}}, \delta)$ if there exists a functional $\boldsymbol{\Upsilon}_\delta$ such that $\boldsymbol{\Upsilon}_\delta \not\equiv \mathbf{0}$ and for all $\mathbf{y_1}, \mathbf{y_2} \in \mathcal{B}(\mathbf{0}_{\mathbf{Y}}, \delta)$ the following inequality holds:

$$\mathbf{J}\left(\frac{\mathbf{y_1} + \mathbf{y_2}}{2}\right) + \boldsymbol{\Upsilon}_\delta(\mathbf{y_1} - \mathbf{y_2}) \leq \frac{\mathbf{1}}{\mathbf{2}}\left(\mathbf{J}(\mathbf{y_1}) + \mathbf{J}(\mathbf{y_2})\right). \tag{315}$$

The functional $\boldsymbol{\Upsilon}_\delta$ enforces standard convexity inequality. For this reason, it is called a **forcing** functional.

It is clear that any uniformly convex functional is convex in $\mathcal{B}(0_{\mathbf{Y}}, \delta)$. Now we establish two important inequalities that hold for uniformly convex functionals.

Theorem

If $\mathbf{J} : \mathbf{Y} \to \overline{\mathbb{R}}$ *is uniformly convex in* $\mathcal{B}(0_{\mathbf{Y}}, \delta)$ *and Gâteaux differentiable in* $\mathcal{B}(0_{\mathbf{Y}}, \delta)$*, then for any* $\mathbf{y}, \mathbf{z} \in \mathcal{B}(\mathbf{0}_{\mathbf{Y}}, \delta)$ *the following relations hold:*

$$\mathbf{J}(\mathbf{z}) \geq \mathbf{J}(\mathbf{y}) + \langle \mathbf{J}'(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + 2\boldsymbol{\Upsilon}_{\delta}(\mathbf{z} - \mathbf{y})$$

and

$$\langle \mathbf{J}'(\mathbf{z}) - \mathbf{J}'(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle \geq 2\boldsymbol{\Upsilon}_{\delta}(\mathbf{z} - \mathbf{y}) + 2\boldsymbol{\Upsilon}_{\delta}(\mathbf{y} - \mathbf{z}).$$

Proof.

We have $\Upsilon_\delta(\mathsf{z}-\mathsf{y}) \le \frac{1}{2}\mathsf{J}(\mathsf{z}) + \frac{1}{2}\mathsf{J}(\mathsf{y}) - \mathsf{J}\left(\frac{\mathsf{z}+\mathsf{y}}{2}\right)$.

Since J is convex and differentiable

$$\mathsf{J}\left(\frac{\mathsf{z}+\mathsf{y}}{2}\right) = \mathsf{J}\left(\mathsf{y}+\frac{\mathsf{z}-\mathsf{y}}{2}\right) \ge \mathsf{J}(\mathsf{y}) + \left\langle \mathsf{J}'(\mathsf{y}), \frac{\mathsf{z}-\mathsf{y}}{2}\right\rangle,$$

and, therefore,

$$2\Upsilon_\delta(\mathsf{z}-\mathsf{y}) \le \mathsf{J}(\mathsf{z}) - \mathsf{J}(\mathsf{y}) - \left\langle \mathsf{J}'(\mathsf{y}), \mathsf{z}-\mathsf{y}\right\rangle.$$

We can rewrite it replacing z by y

$$2\Upsilon_\delta(\mathsf{y}-\mathsf{z}) \le \mathsf{J}(\mathsf{y}) - \mathsf{J}(\mathsf{z}) + \left\langle \mathsf{J}'(\mathsf{z}), \mathsf{z}-\mathsf{y}\right\rangle$$

and obtain the second inequality. □

Deviations from the minimizer

Theorem

Let a functional $\mathbf{J}$ *be uniformly convex in* $\mathcal{B}(0_{\mathbf{Y}}, \delta)$ *and* $\mathbf{y_m} \in \mathcal{B}(\mathbf{0_Y}, \delta)$ *be the minimizer of* $\mathbf{J}$.

$$\Upsilon_\delta(\mathbf{z} - \mathbf{y_m}) \leq \frac{\mathbf{1}}{\mathbf{2}} \left(\mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y_m}) \right), \quad \forall \mathbf{z} \in \mathcal{B}(\mathbf{0_Y}, \delta). \tag{316}$$

Proof.

Since $\mathbf{J}\left(\frac{\mathbf{y_m}+\mathbf{z}}{\mathbf{2}}\right) \geq \mathbf{J}(\mathbf{y_m})$, we obtain

$$\begin{aligned} \Upsilon_\delta(\mathbf{z} - \mathbf{y_m}) \leq \quad & \frac{1}{2}\mathbf{J}(\mathbf{y_m}) + \frac{\mathbf{1}}{\mathbf{2}}\mathbf{J}(\mathbf{z}) - \mathbf{J}\left(\frac{\mathbf{y_m} + \mathbf{z}}{\mathbf{2}}\right) \leq \\ & \leq \frac{1}{2} \left(\mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y_m}) \right). \end{aligned}$$

□

Estimate (316) is the first step in deriving a posteriori error estimates of the functional type by means of the variational techniques. It shows that deviations from the minimizer (measured in terms of the functional Υ_δ) are controlled by the difference of the functionals.

Corollary 1

Rewrite (315) in the form

$$\Upsilon_\delta(z - y_m) + J\left(\frac{y_m + z}{2}\right) - J(y_m) \le \frac{1}{2}\left(J(z) - J(y_m)\right).$$

By virtue of (316), we have

$$J\left(\frac{y_m + z}{2}\right) - J(y_m) \ge 2\Upsilon_\delta\left(\frac{z - y_m}{2}\right)$$

and, therefore, we arrive at the strengthened estimate

$$\Upsilon_\delta(z - y_m) + 2\Upsilon_\delta\left(\frac{z - y_m}{2}\right) \le \frac{1}{2}\left(J(z) - J(y_m)\right). \qquad (317)$$

Corollary 2

Assume that $\mathbf{J}$ is twice differentiable in the vicinity of $\mathbf{y_m}$ and satisfies the finite increment relation

$$\mathbf{J}\left(\frac{\mathbf{y_m}+\mathbf{z}}{2}\right) = \mathbf{J}(\mathbf{y_m}) + \left\langle \mathbf{J}'(\mathbf{y_m}), \frac{\mathbf{z}-\mathbf{y_m}}{2} \right\rangle + \\ + \frac{1}{2}\left\langle \mathbf{J}''\left(\mathbf{y_m} + \xi\frac{\mathbf{z}+\mathbf{y_m}}{2}\right)\frac{\mathbf{z}-\mathbf{y_m}}{2}, \frac{\mathbf{z}-\mathbf{y_m}}{2} \right\rangle,$$

where $\xi \in (0,1)$. Since $\mathbf{J}'(\mathbf{y_m}) = \mathbf{0}_{\mathbf{Y}^*}$, we have another estimate:

$$\Upsilon_\delta(\mathbf{z}-\mathbf{y_m}) + \frac{1}{8}\left\langle \mathbf{J}''\left((1+\tfrac{\xi}{2})\mathbf{y_m} + \tfrac{\xi}{2}\mathbf{z}\right)(\mathbf{z}-\mathbf{y_m}), \mathbf{z}-\mathbf{y_m} \right\rangle \le \\ \le \frac{1}{2}\left(\mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y_m})\right). \quad (318)$$

Example 1

Consider a self-adjoint operator $\mathbf{A} \in \mathcal{L}(\mathbf{H}, \mathbf{H})$ defined on a Hilbert space $\mathbf{H}$ with scalar product $(.,.)$. Assume that it satisfies the condition

$$\alpha_1 \|\mathbf{y}\|^2 \leq \mathbf{G}(\mathbf{y}) := (\mathbf{A}\mathbf{y}, \mathbf{y}) \leq \alpha_2 \|\mathbf{y}\|^2, \qquad \forall \mathbf{y} \in \mathbf{H}.$$

For $\mathbf{J}(\mathbf{y}) = \mathbf{G}(\mathbf{y}) + (\ell, \mathbf{y}), \qquad \ell \in \mathbf{H}$ we have

$$\frac{1}{2}\mathbf{G}(\mathbf{y}) + \frac{1}{2}\mathbf{G}(\mathbf{z}) - \mathbf{G}\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right) =$$

$$= \frac{1}{4}(\mathbf{A}\mathbf{y}, \mathbf{y}) + \frac{1}{4}(\mathbf{A}\mathbf{z}, \mathbf{z}) - \frac{1}{8}(\mathbf{A}(\mathbf{y}+\mathbf{z}), \mathbf{y}+\mathbf{z}) =$$

$$= \frac{1}{8}(\mathbf{A}(\mathbf{z}-\mathbf{y}), (\mathbf{z}-\mathbf{y})),$$

the functional $\mathbf{G}$ is uniformly convex in any ball with

$$\Upsilon(\mathbf{z}-\mathbf{y}) = \frac{1}{8}(\mathbf{A}(\mathbf{z}-\mathbf{y}), (\mathbf{z}-\mathbf{y})).$$

Thus, from (316) we have

$$\frac{1}{8}(\mathbf{A}(\mathbf{z}-\mathbf{y_m}),(\mathbf{z}-\mathbf{y_m})) \le \frac{1}{2}\left(\mathbf{J}(\mathbf{z})-\mathbf{J}(\mathbf{y_m})\right), \quad \forall \mathbf{z}$$

However (318) gives a better estimate

$$\frac{1}{2}(\mathbf{A}(\mathbf{z}-\mathbf{y_m}),(\mathbf{z}-\mathbf{y_m})) \le \mathbf{J}(\mathbf{z})-\mathbf{J}(\mathbf{y_m}). \tag{319}$$

Note that for quadratic type functionals this estimate holds as equality. Indeed,

$$\mathbf{J}(\mathbf{z})-\mathbf{J}(\mathbf{y_m}) = (\mathbf{A}\mathbf{y_m}+\ell,\mathbf{z}-\mathbf{y_m}) + \frac{1}{2}(\mathbf{A}(\mathbf{z}-\mathbf{y_m}),\mathbf{z}-\mathbf{y_m}).$$

and the minimizer $\mathbf{y_m}$ satisfies the relation

$$(\mathbf{A}\mathbf{y_m}+\ell,\mathbf{y}) = \mathbf{0}, \qquad \forall \mathbf{y} \in \mathbf{Y}.$$

Therefore, (319) **holds as equality.**

Theorem

Let $\mathbf{J_1}$ and $\mathbf{J_2}$ be uniformly convex in $\mathcal{B}(0_{\mathbf{Y}}, \delta)$ with functionals $\boldsymbol{\Upsilon}_{1\delta}$ and $\boldsymbol{\Upsilon}_{2\delta}$, respectively.
Then the functional

$$\mu_1 \mathbf{J_1} + \mu_2 \mathbf{J_2},$$

where $\mu_1, \mu_2 \geq 0$, is uniformly convex in $\mathcal{B}(0_{\mathbf{Y}}, \delta)$ with

$$\boldsymbol{\Upsilon}_{\delta} = \mu_1 \boldsymbol{\Upsilon}_{1\delta} + \mu_2 \boldsymbol{\Upsilon}_{2\delta}.$$

Proof.

The proposition follows directly from definition of uniform convexity . □

Example 2

Consider the functional

$$\mathbf{J}(\mathbf{y}) = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{A}\mathbf{y}, \mathbf{y}) + (\ell, \mathbf{y}) + \mathbf{\Psi}(\mathbf{y}),$$

where $\Psi(y)$ is a convex and l.s.c. functional. Applying the above Theorem with $\mu_1 = \mu_2 = 1$,

$$\mathbf{J_1}(\mathbf{y}) = \frac{\mathbf{1}}{\mathbf{2}}(\mathbf{A}\mathbf{y}, \mathbf{y}) + (\ell, \mathbf{y}) \qquad \mathbf{J_2}(\mathbf{y}) = \mathbf{\Psi}(\mathbf{y}),$$

we see that $\mathbf{J}$ is uniformly convex with functional $\mathbf{\Upsilon}$ defined in Example 1.

Theorem

Let $\mathbf{J}_1$ *and* $\mathbf{J}_2$ *be uniformly convex in* $\mathcal{B}(0_{\mathbf{Y}}, \delta)$ *with functionals* $\boldsymbol{\Upsilon}_{1\delta}$ *and* $\boldsymbol{\Upsilon}_{2\delta}$, *respectively. Then the functional*

$$\mathbf{J}(\mathbf{y}) = \max\{\mathbf{J}_1(\mathbf{y}), \mathbf{J}_2(\mathbf{y})\}$$

is uniformly convex in $\mathcal{B}(0_{\mathbf{Y}}, \delta)$ *with*

$$\boldsymbol{\Upsilon}_{\delta} = \min\{\boldsymbol{\Upsilon}_{1\delta}, \boldsymbol{\Upsilon}_{2\delta}\}.$$

Proof. We have

$$\frac{1}{2}\mathbf{J}(\mathbf{y}) + \frac{1}{2}\mathbf{J}(\mathbf{z}) - \mathbf{J}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right) = \frac{1}{2}\max\{\mathbf{J_1}(\mathbf{y}), \mathbf{J_2}(\mathbf{y})\} + \\ + \frac{1}{2}\max\{\mathbf{J_1}(\mathbf{z}), \mathbf{J_2}(\mathbf{z})\} - \max\left\{\mathbf{J_1}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right), \mathbf{J_2}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right)\right\}.$$

Assume that

$$\max\left\{\mathbf{J_1}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right), \mathbf{J_2}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right)\right\} = \mathbf{J_1}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right).$$

Then

$$\frac{1}{2}\left(\mathbf{J}(\mathbf{y}) + \mathbf{J}(\mathbf{z})\right) - \mathbf{J}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right) \geq \\ \geq \frac{1}{2}\left(\mathbf{J_1}(\mathbf{y}) + \mathbf{J_1}(\mathbf{z})\right) - \mathbf{J_1}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right) \geq \boldsymbol{\Upsilon}_{1\delta}(\mathbf{z}-\mathbf{y}).$$

If we have an opposite situation, i.e.,

$$\max\left\{\mathbf{J}_1\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right),\mathbf{J}_2\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right)\right\}=\mathbf{J}_2\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right),$$

then

$$\frac{1}{2}\mathbf{J}(\mathbf{y})+\frac{1}{2}\mathbf{J}(\mathbf{z})-\mathbf{J}\left(\tfrac{\mathbf{y}+\mathbf{z}}{2}\right)\geq\boldsymbol{\Upsilon}_{2\delta}(\mathbf{z}-\mathbf{y}).$$

Thus, in both cases the lower bound is given by the functional

$$\min\ \{\boldsymbol{\Upsilon}_{1\delta}(\mathbf{z}\ -\ \mathbf{y}),\boldsymbol{\Upsilon}_{2\delta}(\mathbf{z}\ -\ \mathbf{y})\}.$$

Example 3. Power growth functionals

Let

$$\mathbf{G(y)} = \frac{\mathbf{1}}{\boldsymbol{\alpha}} \int_{\boldsymbol{\Omega}} |\,\mathbf{y}\,|^{\boldsymbol{\alpha}}\ \mathbf{dx} \qquad \mathbf{F(v)} = \int_{\boldsymbol{\Omega}} \mathbf{fvdx},$$

where $\boldsymbol{\alpha} > 1$. Then Problem $\mathcal{P}$ is to minimize the functional

$$\mathbf{J}_{\boldsymbol{\alpha}}(\mathbf{v}) := \int_{\boldsymbol{\Omega}} \left(\frac{\mathbf{1}}{\boldsymbol{\alpha}} |\,\nabla \mathbf{v}\,|^{\boldsymbol{\alpha}} + \mathbf{fv}\right) \mathbf{dx}$$

over the space $\mathbf{V} = \{\mathbf{v} \in \mathbf{H}^{\boldsymbol{\alpha}}(\boldsymbol{\Omega}) \mid \mathbf{v} = \mathbf{0}\ \text{on}\ \partial\boldsymbol{\Omega}\}$.

Problem $\mathcal{P}^*$ is to maximize the functional

$$\mathbf{I}^*_{\alpha_*}(\mathbf{y}^*) = -\frac{1}{\alpha_*}\int_{\Omega} |\,\mathbf{y}^*\,|^{\alpha^*}\,\mathbf{dx}$$

over the set

$$\mathbf{Q}^*_{\mathbf{f}} = \left\{ \mathbf{y}^* \in \mathbf{Y}^* := \mathbf{L}^{\alpha_*}(\Omega, \mathbb{R}^{\mathbf{n}}) \middle| \int_{\Omega} \mathbf{y}^* \cdot \nabla \mathbf{w} \mathbf{dx} = \int_{\Omega} \mathbf{f} \mathbf{w} \mathbf{dx}\ \forall \mathbf{w} \in \mathbf{V} \right\}.$$

For $\alpha \geq 2$ uniform convexity of $\mathbf{G(y)}$ follows from the first Clarkson's inequality

$$\int_{\Omega} \left| \frac{y_1+y_2}{2} \right|^{\alpha} dx + \int_{\Omega} \left| \frac{y_1-y_2}{2} \right|^{\alpha} dx \leq \frac{1}{2} \int_{\Omega} (|y_1|^{\alpha} + |y_2|^{\alpha})\, dx,$$

which is valid for all $y_1, y_2 \in Y$.

See S. L. Sobolev. *Some Applications of Functional Analysis in Mathematical Physics*. Hence, we observe that in this case

$$\Upsilon_{\ominus}(z) = \frac{1}{\alpha} \|z\|^{\alpha}_{\alpha,\Omega}.$$

and

$$\frac{1}{\alpha} \int_{\Omega} |\nabla(v-u)|^{\alpha} dx \leq \frac{1}{\alpha} (J_{\alpha}(v) - I^{*}_{\alpha}(q^{*})), \ \forall q^{*} \in Q^{*}_{f},$$

For $1 < \alpha \leq 2$, the functional $\mathbf{G}$ is also uniformly convex. This fact follows from the second Clarkson's inequality

$$\left(\int_\Omega \left(\frac{\mathbf{y}_1 + \mathbf{y}_2}{2} \right)^\alpha \mathbf{dx} \right)^{\frac{1}{\alpha-1}} + \left(\int_\Omega \left(\frac{\mathbf{y}_1 - \mathbf{y}_2}{2} \right)^\alpha \mathbf{dx} \right)^{\frac{1}{\alpha-1}} \leq \left(\frac{1}{2} \int_\Omega \left(|\, \mathbf{y}_1 \,|^\alpha + |\, \mathbf{y}_2 \,|^\alpha \right) \mathbf{dx} \right)^{\frac{1}{\alpha-1}} .$$

However, in this case, the functional $\boldsymbol{\Upsilon}_\delta$ depends on the radius δ of a ball $\mathfrak{B}(\mathbf{0_Y}, \delta)$ that contains $\mathbf{y_1}$ and $\mathbf{y_2}$, so that the estimate holds with

$$\boldsymbol{\Upsilon}_\delta(\mathbf{z}) = \delta^{\frac{\alpha-2}{\alpha-1}} \kappa \, \|\mathbf{z}\|_{\boldsymbol{\alpha},\Omega}^{\frac{\alpha}{\alpha-1}} ,$$

where $\kappa = \frac{1}{\kappa_0+1}$ and κ_0 is the integer part of $\frac{1}{\alpha-1}$.

Now we introduce a general scheme for deriving a posteriori error estimates by using duality theory of the calculus of variations. We consider variational problems of the form

$$\inf_{\mathbf{v}\in\mathbf{V}}\{\mathbf{F}(\mathbf{v})+\mathbf{G}(\mathbf{\Lambda v})\},$$

where $\mathbf{F}:\mathbf{V}\to\mathbb{R}$ is a convex lower semicontinuous functional, $\mathbf{G}:\mathbf{Y}\to\mathbb{R}$ is a uniformly convex functional, $\mathbf{V}$ and $\mathbf{Y}$ are reflexive Banach spaces and $\mathbf{\Lambda}:V\to Y$ is a bounded linear operator.

General variational problem

Consider the general variational problem: find $\mathbf{u}$ in a Banach space V such that

$$\mathbf{J}(\mathbf{u}, \mathbf{\Lambda u}) = \inf_{\mathbf{v} \in \mathbf{V}} \mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}), \tag{320}$$

where $\mathbf{J}(\mathbf{v}) = \mathbf{F}(\mathbf{v}) + \mathbf{G}(\mathbf{\Lambda v})$, $\mathbf{F}$ is a convex, lower semicontinuous functional, $\mathbf{G}$ is a uniformly convex functional and $\mathbf{\Lambda} : V \longrightarrow Y$ is a bounded linear operator.
$\mathbf{V}$ and $\mathbf{Y}$ are reflexive Banach spaces endowed with the norms $\|.\|_V$ and $\|.\|$, respectively.

Dual spaces are denoted by $\mathbf{V}^*$ and $\mathbf{Y}^*$ with duality pairings $\langle ., . \rangle$ and $\langle\langle ., . \rangle\rangle$, respectively. The spaces $\mathbf{Y}$ and $\mathbf{Y}^*$ are endowed with the norms $\|.\|$ and $\|.\|_*$.
We assume that

$$\|\mathbf{\Lambda w}\| \geq \mathbf{c_0} \|\mathbf{w}\|_{\mathbf{V}} \quad \forall \mathbf{w} \in \mathbf{V}\,, \tag{321}$$

where $\mathbf{c_0}$ is a positive constant independent of $\mathbf{w}$.
In addition to $\mathbf{\Lambda}$, we introduce its conjugate $\mathbf{\Lambda}^* : Y^* \longrightarrow V^*$. This amounts to say that

$$\langle\langle \mathbf{y}^*, \mathbf{\Lambda v} \rangle\rangle = \langle \mathbf{\Lambda}^* \mathbf{y}^*, \mathbf{v} \rangle \;\; \forall \mathbf{y}^* \in \mathbf{Y}^*, \mathbf{v} \in \mathbf{V}\,. \tag{322}$$

$\mathbf{J(v, \Lambda v)} := \mathbf{F(v)} + \mathbf{G(\Lambda v)}$. is assumed to be coercive on $\mathbf{V}$, i.e.

$$\mathbf{J(v, \Lambda v)} \longrightarrow +\infty \quad \text{if} \quad \|\mathbf{v}\|_{\mathbf{V}} \longrightarrow +\infty.$$

Primal and Dual Problems

Problem $\mathcal{P}$**.** *Find* $\mathbf{u} \in \mathbf{V}$ *such that*

$$\mathbf{J}(\mathbf{u}, \mathbf{\Lambda u}) = \mathbf{inf}\, \mathcal{P} := \inf_{\mathbf{v} \in \mathbf{V}} \mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}). \tag{323}$$

The problem dual to (323 is (see e.g.
I. Ekeland and R. Temam *Convex analysis and variational problems.* North-Holland, Amsterdam, 1976.)

Problem $\mathcal{P}^*$**.** *Find* $\mathbf{p}^* \in \mathbf{Y}^*$ *such that*

$$\begin{aligned} - \quad & \mathbf{J}^*(\mathbf{\Lambda}^* \mathbf{p}^*, -\mathbf{p}^*) = \mathbf{sup}\, \mathcal{P}^* := \sup_{\mathbf{y}^* \in \mathbf{Y}^*} -\mathbf{J}^*(\mathbf{\Lambda}^* \mathbf{y}^*, -\mathbf{y}^*), \\ & \mathbf{J}^*(\mathbf{\Lambda}^* \mathbf{y}^*, -\mathbf{y}^*) := \mathbf{F}^*(\mathbf{\Lambda}^* \mathbf{y}^*) + \mathbf{G}^*(-\mathbf{y}^*), \end{aligned} \tag{324}$$

where $\mathbf{F}^*$ *and* $\mathbf{G}^*$ *are the functionals conjugate of* $\mathbf{F}$ *and* $\mathbf{G}$*, respectively.*

Theorem (1)

If the functional $\mathbf{F}$ *is finite at some* $\mathbf{u_0} \in \mathbf{V}$ *and the functional* $\mathbf{G}$ *is continuous and finite at* $\mathbf{\Lambda u_0} \in \mathbf{Y}$*, then there exists a minimizer* $\mathbf{u}$ *to Problem* $\mathcal{P}$ *and a maximizer* $\mathbf{p}^*$ *to Problem* $\mathcal{P}^*$*. Besides,*

$$\mathbf{inf}\,\mathcal{P} = \mathbf{sup}\,\mathcal{P}^* \tag{325}$$

and the following duality relations hold

$$\begin{aligned} &\textbf{(i)} \quad \mathbf{F(u)} + \mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{p}^*) - \langle \mathbf{\Lambda}^*\mathbf{p}^*, \mathbf{u}\rangle = \mathbf{0}, \\ &\textbf{(ii)} \quad \mathbf{G(\Lambda u)} + \mathbf{G}^*(-\mathbf{p}^*) + \langle\langle \mathbf{p}^*, \mathbf{\Lambda u}\rangle\rangle = \mathbf{0}. \end{aligned} \tag{326}$$

Above relations are equivalent to

$$\textbf{(i)}\ \mathbf{\Lambda}^*\mathbf{p}^* \in \partial\mathbf{F(u)}, \qquad \textbf{(ii)}\ -\mathbf{p}^* \in \partial\mathbf{G(\Lambda u)}.$$

Problems with uniformly convex functionals

We recall (see Lecture 4) that a continuous functional $\mathbf{G} : \mathbf{Y} \longrightarrow \mathbb{R}$ is uniformly convex in a ball $\mathbf{B}(\mathbf{0}, \delta) := \{\mathbf{y} \in \mathbf{Y} \mid\mid \|\mathbf{y}\| < \delta\}$ if there exists a continuous functional $\mathbf{\Phi}_\delta : \mathbf{Y} \longrightarrow \mathbb{R}_+$ such that $\mathbf{\Phi}_\delta(\mathbf{y}) = \mathbf{0}$ only if $\mathbf{y} = \mathbf{O_y}$ is and

$$\mathbf{G}(\tfrac{\mathbf{y_1}+\mathbf{y_2}}{\mathbf{2}}) + \mathbf{\Phi}_\delta(\mathbf{y_2} - \mathbf{y_1}) \leq \tfrac{\mathbf{1}}{\mathbf{2}} (\mathbf{G}(\mathbf{y_1}) + \mathbf{G}(\mathbf{y_2})) \quad \forall\, \mathbf{y_1}, \mathbf{y_2} \in \mathbf{B}(\mathbf{0}, \delta).$$

Usually, $\mathbf{\Phi}_\delta$ is given by a continuous strictly increasing function of the norm $\|\mathbf{y}\|$.

General form of a posteriori estimates for uniformly convex variational problems was established in

S. Repin. A posteriori error estimation for variational problems with uniformly convex functionals, *Math. Comput.*, 69(230), 2000, 481-500.

General form of the functional a posteriori estimate

Theorem (2)

Assume that the above conditions on $\mathbf{F}$ *and* $\mathbf{G}$ *are satisfied and*
(i) $\mathbf{G}$ *is uniformly convex on a ball* $B(0,\delta)$,
(ii) the solution $\mathbf{u}$ *of Problem* $\mathcal{P}$ *and an element* $\mathbf{v} \in \mathbf{V}$ *are such, that* $\mathbf{\Lambda u}, \mathbf{\Lambda v} \in \mathbf{B}(\mathbf{0},\delta)$.
Then, for any $\mathbf{y}^* \in \mathbf{Y}^*$

$$\mathbf{\Phi}_\delta\left(\mathbf{\Lambda}(\mathbf{v}-\mathbf{u})\right) \leq \mathbf{M}_\oplus\left(\mathbf{v},\mathbf{y}^*\right) := \mathbf{D_F}(\mathbf{\Lambda}^*\mathbf{y}^*,\mathbf{v}) + \mathbf{D_G}(\mathbf{y}^*,\mathbf{\Lambda v}) \qquad (327)$$

where

$$\begin{aligned} \mathbf{D_F}(\mathbf{\Lambda}^*\mathbf{y}^*,\mathbf{v}) &:= \tfrac{1}{2}\left(\mathbf{F}(\mathbf{v}) + \mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) - \langle \mathbf{\Lambda}^*\mathbf{y}^*,\mathbf{v}\rangle\right), \\ \mathbf{D_G}(\mathbf{y}^*,\mathbf{\Lambda v}) &:= \tfrac{1}{2}\left(\mathbf{G}(\mathbf{\Lambda v}) + \mathbf{G}^*(-\mathbf{y}^*) + \langle\langle \mathbf{y}^*,\mathbf{\Lambda v}\rangle\rangle\right). \end{aligned}$$

Proof

Since $\mathbf{F}$ is convex and $\mathbf{G}$ is uniformly convex we obtain

$$\mathbf{\Phi}_\delta\left(\mathbf{\Lambda}(\mathbf{v}-\mathbf{u})\right)+\mathbf{G}(\mathbf{\Lambda}(\tfrac{\mathbf{v}+\mathbf{u}}{2}))+\mathbf{F}(\tfrac{\mathbf{v}+\mathbf{u}}{2})\leq \tfrac{1}{2}\Big[(\mathbf{F}(\mathbf{v})+\mathbf{G}(\mathbf{\Lambda v}))+(\mathbf{F}(\mathbf{u})+\mathbf{G}(\mathbf{\Lambda u}))\Big].$$

The element $\mathbf{u}$ is a minimizer, therefore

$$\mathbf{G}(\mathbf{\Lambda u})+\mathbf{F}(\mathbf{u})=\mathbf{J}(\mathbf{u})\leq \mathbf{G}(\mathbf{\Lambda}\left(\tfrac{\mathbf{u}+\mathbf{v}}{2}\right))+\mathbf{F}(\tfrac{\mathbf{u}+\mathbf{v}}{2})$$

and we have

$$\mathbf{\Phi}_\delta\left(\mathbf{\Lambda}(\mathbf{v}-\mathbf{u})\right)+\mathbf{G}(\mathbf{\Lambda u})+\mathbf{F}(\mathbf{u})\leq \tfrac{1}{2}\Big[(\mathbf{F}(\mathbf{v})+\mathbf{G}(\mathbf{\Lambda v}))+(\mathbf{F}(\mathbf{u})+\mathbf{G}(\mathbf{\Lambda u}))\Big].$$

From the above we observe that

$$\Phi_\delta(\Lambda e) \le \tfrac{1}{2}\Big[(F(v)+G(\Lambda v))-(F(u)+G(\Lambda u))\Big] = \\ = \tfrac{1}{2}(J(v,\Lambda v)-J(u,\Lambda u)) \quad \forall v \in B(0,\delta)\,.$$

In view of Theorem 1,

$$J(u,\Lambda u) = \inf \mathcal{P} = \sup \mathcal{P}^* = -F^*(\Lambda^* p^*) - G^*(-p^*).$$

Since p^* is a solution of the dual problem, we have

$$-J^*(\Lambda^* p^*, -p^*) \ge -J^*(\Lambda^* y^*, -y^*) \quad \forall y^* \in Y^*,$$

so that

$$J(u,\Lambda u) \ge -F^*(\Lambda^* y^*) - G^*(-y^*).$$

Therefore

$$\begin{aligned} \mathbf{\Phi}_\delta(\mathbf{\Lambda e}) &\leq \tfrac{1}{2}\left(\mathbf{F(v)} + \mathbf{G(\Lambda v)} + \mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{p}^*) + \mathbf{G}^*(-\mathbf{p}^*)\right) \leq \\ &\leq \tfrac{1}{2}\left(\mathbf{F(v)} + \mathbf{G(\Lambda v)} + \mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) + \mathbf{G}^*(-\mathbf{y}^*)\right) . \end{aligned}$$

However, by (322) we observe that

$$\langle\langle \mathbf{y}^*, \mathbf{\Lambda v}\rangle\rangle - \langle \mathbf{\Lambda}^*\mathbf{y}^*, \mathbf{v}\rangle = \mathbf{0} \;\; \forall \mathbf{y}^* \in \mathbf{Y}^*, \, \mathbf{v} \in \mathbf{V} .$$

We add this zero term to the above relation and obtain the required estimate.

□

Comments

The right–hand side of (327 is the **sum of two compound functionals**

$$\mathbf{M_F} : \mathbf{V}^* \times \mathbf{V} \longrightarrow \mathbb{R} \quad \text{and} \quad \mathbf{M_G} : \mathbf{Y}^* \times \mathbf{Y} \longrightarrow \mathbb{R}.$$

They are nonnegative and vanishes if and only if $\mathbf{v}$ and $\mathbf{y}^*$ satisfy the relations (326)(i)–(ii).

Therefore, $\mathbf{M}_\oplus(\mathbf{v}, \mathbf{y}^*)$ is, in fact, a measure of the error in the **duality relations** for the pair $(\mathbf{v}, \mathbf{y}^*)$.
It vanishes if and only if $\mathbf{v} = \mathbf{u}$ and $\mathbf{y}^* = \mathbf{p}^*$.

Let the functional $\mathbf{F}$ be uniformly convex on $\mathbf{V}$ with a forcing functional φ_δ. Then the "forcing functional" has the form we have

$$\mathbf{\Phi}_\delta(\mathbf{\Lambda e}) + \varphi_\delta(\mathbf{e}) \quad \leq \tfrac{1}{2}(\mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}) - \mathbf{J}(\mathbf{u}, \mathbf{\Lambda u})) \tag{328}$$

and, as a result, (327) is replaced by the strengthened estimate

$$\mathbf{\Phi}_\delta(\mathbf{\Lambda e}) + \varphi_\delta(\mathbf{e}) \leq \mathbf{M}_\oplus(\mathbf{v}, \mathbf{y}^*) \quad \forall \mathbf{y}^* \in \mathbf{Y}^* . \tag{329}$$

It is not difficult to verify that

$$\begin{aligned}
&\mathsf{M}_{\oplus}(\mathbf{v},\mathbf{y}^*) - \mathsf{M}_{\oplus}(\mathbf{v},\mathbf{p}^*) = \\
&= \tfrac{1}{2}\left(\mathsf{F}(\mathbf{v})+\mathsf{F}^*(\boldsymbol{\Lambda}^*\mathbf{y}^*)-\langle\boldsymbol{\Lambda}^*\mathbf{y}^*,\mathbf{v}\rangle+\mathsf{G}(\boldsymbol{\Lambda}\mathbf{v})+\mathsf{G}^*(-\mathbf{y}^*)+\langle\!\langle\mathbf{y}^*,\boldsymbol{\Lambda}\mathbf{v}\rangle\!\rangle\right) - \\
&\tfrac{1}{2}\left(\mathsf{F}(\mathbf{v})+\mathsf{F}^*(\boldsymbol{\Lambda}^*\mathbf{p}^*)-\langle\boldsymbol{\Lambda}^*\mathbf{p}^*,\mathbf{v}\rangle+\mathsf{G}(\boldsymbol{\Lambda}\mathbf{v})+\mathsf{G}^*(-\mathbf{p}^*)+\langle\!\langle\mathbf{p}^*,\boldsymbol{\Lambda}\mathbf{v}\rangle\!\rangle\right) = \\
&\qquad = \mathsf{J}^*(\boldsymbol{\Lambda}^*\mathbf{y}^*,-\mathbf{y}^*) - \mathsf{J}^*(\boldsymbol{\Lambda}^*\mathbf{p}^*,-\mathbf{p}^*) =\geq \mathbf{0}.
\end{aligned}$$

Therefore, for any $\mathbf{v}$ the right-hand side of (327) is minimal if $\mathbf{y}^* = \mathbf{p}^*$. Consequently, to make the estimate effective we have to find some $\mathbf{y}^*$ close to $\mathbf{p}^*$ in $\mathbf{Y}^*$. A simple way to obtain a function "close" to $\mathbf{p}^*$ it to use duality relations. To this end, we set $\mathbf{y}^* = \boldsymbol{\sigma}^*(\mathbf{v})$, where

$$-\boldsymbol{\sigma}^*(\mathbf{v}) \in \partial\mathsf{G}(\boldsymbol{\Lambda}\mathbf{v}) \subset \mathbf{Y}^*.$$

In this case,

$$\mathsf{M}_{\mathsf{G}}(\boldsymbol{\sigma}^*(\mathbf{v}), \boldsymbol{\Lambda}\mathbf{v}) = \mathbf{0}$$

and we get the estimate

$$\boldsymbol{\Phi}_{\delta}\left(\boldsymbol{\Lambda}\mathbf{e}\right) \leq \mathsf{M}_{\mathsf{F}}(\boldsymbol{\Lambda}^*\boldsymbol{\sigma}^*(\mathbf{v}), \mathbf{v}) \tag{330}$$

whose right–hand side depends on $\mathbf{v}$ only.
However, the estimate (330) cannot be directly applied in one practically important case which we consider below.

Example. Diffusion problem with Robin conditions

Consider the variational problems for the functional

$$\mathbf{J}(\mathbf{v}, \nabla \mathbf{v}) = \int_{\Omega} (\frac{\mathbf{1}}{\mathbf{2}} |\nabla \mathbf{v}|^2 + \frac{\delta}{\mathbf{2}} |\mathbf{v}|^2) \mathbf{dx} + \int_{\partial_2 \Omega} (\frac{\alpha}{\mathbf{2}} |\mathbf{v}|^2 - \mathbf{gv}) \, \mathbf{ds} \, .$$

Our problem is to minimize $\mathbf{J}$ on the set of functions vanishing at $\partial_1 \Omega$. Minimizer $\mathbf{u}$ of this variational problem is related to the system

$$\begin{aligned} -\mathbf{\Delta u} + \delta \mathbf{u} &= \mathbf{0}, && \text{in } \mathbf{\Omega}, \\ \frac{\partial \mathbf{u}}{\partial \mathbf{n}} + \alpha \mathbf{u} - \mathbf{g} &= \mathbf{0}, && \text{on } \partial_2 \mathbf{\Omega}. \end{aligned}$$

On $\partial_2 \Omega$ the solution satisfies the so–called Robin boundary condition. Let us show that the respective functional a posteriori estimate for the problem with Robin type boundary conditions easily follows from the above general estimate.

We set

$$\mathbf{\Lambda v} := \nabla \mathbf{v},$$

$$\mathbf{G}(\mathbf{\Lambda w}) = \int_{\Omega} \frac{\mathbf{1}}{\mathbf{2}} |\nabla \mathbf{v}|^2 \mathbf{dx}$$

and

$$\mathbf{F}(\mathbf{v}) = \int_{\Omega} \frac{\delta}{\mathbf{2}} |\mathbf{v}|^2 \mathbf{dx} + \int_{\partial_2 \Omega} (\frac{\alpha}{\mathbf{2}} |\mathbf{v}|^2 - \mathbf{gv}) \, \mathbf{ds} \, .$$

Since

$$\int_{\Omega} \mathbf{y}^* \cdot \nabla \mathbf{v} \mathbf{dx} = \int_{\Omega} -\mathbf{div} \mathbf{y}^* \mathbf{v} \mathbf{dx} + \int_{\partial_2 \Omega} (\mathbf{y}^* \mathbf{n}) \mathbf{v} \mathbf{ds},$$

we observe that $\mathbf{\Lambda}^* \mathbf{y}^* = \{-\mathbf{div} \mathbf{y}^* \mid_{\Omega}, \mathbf{y}^* \cdot \mathbf{n} \mid_{\partial_2 \Omega}\}$.
In the considered case,

$$\langle \mathbf{y}^*, \mathbf{y} \rangle := \int_{\Omega} \mathbf{y}^* \cdot \mathbf{y} \, \mathbf{dx};$$

$$\mathbf{G}^*(-\mathbf{y}^*) = \sup_{\mathbf{y}} \int_{\Omega} (-\mathbf{y}^* \cdot \mathbf{y} - \frac{\mathbf{1}}{\mathbf{2}} |\mathbf{y}|^2) \mathbf{dx} = \int_{\Omega} \frac{\mathbf{1}}{\mathbf{2}} |\mathbf{y}^*|^2 \mathbf{dx}.$$

Therefore,

$$\mathbf{G}(\mathbf{\Lambda v}) + \mathbf{G}^*(-\mathbf{y}^*) + \langle \mathbf{y}^*, \mathbf{\Lambda v} \rangle = \int_{\Omega} (\frac{\mathbf{1}}{\mathbf{2}} |\nabla \mathbf{v}|^2 + \frac{\mathbf{1}}{\mathbf{2}} |\mathbf{y}^*|^2 + \nabla \mathbf{v} \cdot \mathbf{y}^*) \mathbf{dx}.$$

Next, in general,

$$\langle\langle \mathbf{\Lambda}^* \mathbf{y}^*, \mathbf{v} \rangle\rangle = < -\mathbf{div} \mathbf{y}^*, \mathbf{v} >_{\mathbf{H}^{-1}(\Omega)} + < \mathbf{y}^* \cdot \mathbf{n}, \mathbf{v} >_{\mathbf{H}^{-1/2}(\partial_2 \Omega)} .$$

However, if we assume that y^* is sufficiently regular, then

$$\langle\langle \mathbf{\Lambda}^* \mathbf{y}^*, \mathbf{v} \rangle\rangle = \int_{\Omega} -\mathbf{div} \mathbf{y}^* \mathbf{v} \mathbf{dx} + \int_{\partial_2 \Omega} \mathbf{y}^* \cdot \mathbf{n} \mathbf{v} \mathbf{ds}.$$

Now,

$$\mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) = \sup_{\mathbf{v}}\{\int_{\Omega} -\mathbf{div}\mathbf{y}^*\mathbf{v}\mathbf{dx} + \int_{\partial_2\Omega} \mathbf{y}^*\cdot\mathbf{n}\mathbf{v}\mathbf{ds} - \mathbf{F}(\mathbf{v})\} =$$

$$\sup_{\mathbf{v}}\{\int_{\Omega} -\mathbf{div}\mathbf{y}^*\mathbf{v}\mathbf{dx} + \int_{\partial_2\Omega} \mathbf{y}^*\cdot\mathbf{n}\mathbf{v}\mathbf{ds} - \int_{\Omega} \frac{\delta}{2}|\mathbf{v}|^2\mathbf{dx} - \int_{\partial_2\Omega} (\frac{\alpha}{2}|\mathbf{v}|^2 - \mathbf{g}\mathbf{v})\,\mathbf{ds} \leq$$

$$\sup_{\mathbf{v}\in\mathbf{L}_2(\Omega)} \int_{\Omega} (-\mathbf{div}\mathbf{y}^*\mathbf{v} - \frac{\delta}{2}|\mathbf{v}|^2)\mathbf{dx} + \sup_{\varrho\in\mathbf{L}_2(\partial_2\Omega)} \int_{\partial_2\Omega} ((\mathbf{y}^*\cdot\mathbf{n})\varrho - \frac{\alpha}{2}|\varrho|^2 + \mathbf{g}\varrho)\mathbf{ds}$$

$$\sup_{\mathbf{v}\in \mathbf{L}_2(\Omega)} \int_{\Omega} (-\mathbf{div}\mathbf{y}^*\mathbf{v} - \frac{\delta}{2}|\mathbf{v}|^2)\mathbf{dx} = \int_{\Omega} \frac{\mathbf{1}}{2\delta}|\mathbf{div}\mathbf{y}^*|^2\mathbf{dx},$$

$$\sup_{\varrho\in \mathbf{L}_2(\partial_2\Omega)} \int_{\partial_2\Omega} ((\mathbf{y}^*\cdot\mathbf{n})\varrho - \frac{\alpha}{2}|\varrho|^2 + \mathbf{g}\varrho)\mathbf{ds} = \int_{\partial_2\Omega} \frac{\mathbf{1}}{2\alpha}|\mathbf{y}^*\cdot\mathbf{n}+\mathbf{g}|^2\mathbf{ds}.$$

Hence,

$$\mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) \le \int_{\Omega} \frac{\mathbf{1}}{2\delta}|\mathbf{div}\mathbf{y}^*|^2\mathbf{dx} + \int_{\partial_2\Omega} \frac{\mathbf{1}}{2\alpha}|\mathbf{y}^*\cdot\mathbf{n}+\mathbf{g}|^2\mathbf{ds}.$$

Now,

$$\mathbf{F(v)} = \int_{\Omega} \frac{\delta}{2}|\mathbf{v}|^2 \mathbf{dx} + \int_{\partial_2\Omega} (\frac{\alpha}{2}|\mathbf{v}|^2 - \mathbf{gv})\,\mathbf{ds}\,,$$

$$\langle\langle \mathbf{\Lambda^* y^*, v} \rangle\rangle = \int_{\Omega} -\mathbf{div y^* v dx} + \int_{\partial_2\Omega} (\mathbf{y^* \cdot n}) \mathbf{v ds},$$

$$\mathbf{F^*(\Lambda^* y^*)} \leq \int_{\Omega} \frac{\mathbf{1}}{\mathbf{2}\delta}|\mathbf{div y^*}|^2 \mathbf{dx} + \int_{\partial_2\Omega} \frac{\mathbf{1}}{\mathbf{2}\alpha}|\mathbf{y^* \cdot n + g}|^2 \mathbf{ds}.$$

Therefore,

$$\mathbf{F(v)} + \mathbf{F^*(\Lambda^* y^*)} - \langle\langle \mathbf{\Lambda^* y^*, v} \rangle\rangle \leq \int_{\Omega} \frac{\mathbf{1}}{\mathbf{2}\delta}(\mathbf{div y^*} + \delta \mathbf{v})^2 \mathbf{dx} + \int_{\partial_2\Omega} \left(\frac{\alpha}{2}|\mathbf{v}|^2 + \frac{\mathbf{1}}{\mathbf{2}\alpha}|\mathbf{y^* \cdot n + g}|^2 - (\mathbf{y^* \cdot n + g})\mathbf{v}\right) \mathbf{ds}.$$

We obtain

$$\mathbf{F}(\mathbf{v}) + \mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) - \langle\langle \mathbf{\Lambda}^*\mathbf{y}^*, \mathbf{v} \rangle\rangle \leq \int_{\Omega} \frac{\mathbf{1}}{\mathbf{2}\delta}(\mathbf{div}\mathbf{y}^* + \delta\mathbf{v})^2 \mathbf{dx} + \int_{\partial_2\Omega} \frac{\mathbf{1}}{\mathbf{2}\alpha}|\mathbf{y}^* \cdot \mathbf{n} + \mathbf{g} - \alpha\mathbf{v}|^2 \mathbf{ds}.$$

$$\mathbf{G}(\mathbf{\Lambda}\mathbf{v}) + \mathbf{G}^*(-\mathbf{y}^*) + \langle \mathbf{y}^*, \mathbf{\Lambda}\mathbf{v} \rangle = \int_{\Omega} \frac{\mathbf{1}}{\mathbf{2}}|\nabla\mathbf{v} + \mathbf{y}^*|^2 \mathbf{dx}.$$

Two terms above give the error Majorant.

We observe that the Majorant vanishes if and only if

$$\begin{aligned} &\mathbf{div}\,\mathbf{y}^* + \delta\mathbf{v} = \mathbf{0} \qquad \text{in } \boldsymbol{\Omega},\\ &\mathbf{y}^* \cdot \mathbf{n} + \mathbf{g} - \alpha\mathbf{v} = \mathbf{0} \qquad \text{on } \partial_2\boldsymbol{\Omega},\\ &\mathbf{y}^* = -\nabla\mathbf{v} \qquad \text{in } \boldsymbol{\Omega}. \end{aligned}$$

These relations mean that

$$\begin{aligned} &-\boldsymbol{\Delta}\mathbf{v} + \delta\mathbf{v} = \mathbf{0} \qquad \text{in } \boldsymbol{\Omega},\\ &\frac{\partial\mathbf{v}}{\partial\mathbf{n}} + \alpha\mathbf{v} = \mathbf{g} \qquad \text{on } \partial_2\boldsymbol{\Omega}, \end{aligned}$$

i.e., since $\mathbf{v}$ vanishes at $\partial_1\Omega$ it is but the exact solution.


M. Ainsworth and J. T. Oden. A posteriori error estimation in the finite element method, *Numer. Math.*, 60(1992) 429-463.


M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods, *Numer. Math.*, 65(1993) 23-50.


M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Wiley and Sons, New York, 2000.


M. Ainsworth, J. T. Oden and C. Y. Lee. Local a posteriori error estimators for variational inequalities, *Numer. Methods for PDE*, 9(1993), 23-33.


M. Ainsworth, J. Z. Zhu, A. W. Craig and O. C. Zienkiewicz. Analysis of the Zienkiewicz-Zhu a posteriori error estimator in the finite element method, *Int. J. Numer. Methods Engrg.*, 28(1989), 2161-2174.


M. Amara, M. Ben Younes and C. Bernardi. Error indicators for the Navier-Stokes equations in stream function and vorticity formulation, *Numer. Math.*, 80(1998), 181-206.

A. Arkhipova. On the best possible smoothness of the problem with two-side constraints, *Vestnik Leningr. Univ., ser. Math.*, 7 (1984), 5-9 (in Russian).

A. M. Arthurth. Complementary variational principles, Clarendon Press, Oxford, 1980.

G. Astarita and G. Marrucci. *Principles of non Newtonian fluid mechanics*. McGraw Hill, London, 1974.

J.P.Aubin. *Approximation of elliptic boundary value problems*. Wiley-Interscience, New York-London-Sydney, 1972.

G. Auchmuty. A posteriori error estimates for linear equations, *Numer. Math.*, 61(1992), 1-6.

O. Axelsson. *Iterative solution methods*. Cambridge University Press, Cambridge, 1994.

I. Babuška. Courant element: before and after, in *Fifty years of Courant element*, M. Křižek, P. Neittaanmäki and R. Stenberg (Eds.), Marcel Dekker, 1994, 37-51.


I. Babuška. The finite element method with Lagrangian multipliers, *Numer. Math.*, 20(1973), 179-192.


I. Babuška, R. Duran and R. Rodriguez. Analysis of the efficiency of a posteriori error estimator for linear triangular elements, *SIAM J. Numer. Anal.* 29(1992), 947-964.


I. Babuška, F. Ihlenburg, T. Strouboulis and S. K. Gangaraj. A posteriori error estimation for finite element solutions on Helmholtz' equation–Part II: estimation of the pollution error, *Int. J. Numer. Meth. Engrg.*, 40(1997), 3883-3900.


I. Babuška and J. E. Osborn. Can a finite element method perform arbitrarily badly?, *Math. Comput.*, 69(2000), 230, 443-462.


I. Babuška and W. C. Rheinboldt. A-posteriori error estimates for the finite element method. Internat. J. Numer. Meth. Engrg., 12(1978) 1597-1615.


I. Babuška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. SIAM J.Numer. Anal., 15(1978), 736-754.


I. Babuška and R. Rodriguez. The problem of the selection of an a posteriori error indicator based on smoothing techniques, *Internat. J. Numer. Meth. Engrg.*, 36(1993), 539-567.


I. Babuška and T. Strouboulis. *The finite element method and its reliability*. The Clarendon Press, Oxford University Press, New York, 2001.


I. Babuška, T. Strouboulis, C. S. Upadhyay and S. K. Gangaraj. A posteriori estimation and adaptive control of the pollution-error in the *h*-version of the FEM, *Int. J. Numer. Meth. Engrg.*, 38(1995), 4207-4235.


I. Babuška, T. Strouboulis, S. K. Gangaraj and C. S. Upadhyay. Pollution-error in the *h*-version of the FEM and the local quality of recovered derivatives, *Comput. Methods Appl. Mech. Engrg.*,140(1997), 1-37.


I. Babuška, T. Strouboulis, A. Mathur and C. S. Upadhyay. Pollution error in the *h*-version of the FEM and the local quality of a posteriori error estimators, *Finite. Elem. Anal. and Des.*, 17(1994), 273-321.

I. Babuška, T. Strouboulis and C. S. Upadhyay. A model study of the quality of a posteriori error estimators for linear elliptic problems. Error estimation in the interior of patchwise uniform grids of triangles, *Comp. Meth. Appl. Mech. Engrg.* , 114(1994), 307-378.

C. Baiocchi and A. Capelo. *Variational and quasivariational inequalities. Applications to free boundary problems.* Wiley and Sons, New York, 1984.

J. M. Ball and B. D. James. Fine phase mixtures as minimizers of energy, *Arch. Rational Mech. Anal.*, 100(1987), 13-52.

P. Baumann and D. Phillips. A nonconvex variational problem related to change of phase, *Appl. Math. Optim.*, 21(1990), 113-138.

R. E. Bank and R. K. Smith. Mesh smoothing using a posteriori error estimates, *SIAM J. Numer. Anal.*, 34(1997), 3, 979-997.

R. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations, *Math. Comput.*, 44(1985), 283-301.


R. E. Bank and B. D. Welfert. A posteriori error estimates for the Stokes problem, *SIAM J. Numer. Anal.*, 28(1991), 591-623.


W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations.* Birkhäuser, Berlin, 2003.


R. Becker and R. Rannacher. A feed–back approach to error control in finite element methods: Basic approach and examples, *East–West J. Numer. Math.*, 4(1996), 237-264.


R. Becker, H. Kapp and R. Rannacher. Adaptive finite element methods for optimal control of partial differential equations: Basic concept. *SIAM J. Control Optim.*, 39(2000),1, 113-132.


A. Bermúdez, R. Durán and R. Rodríguez. Finite element analysis of compressible and incompressible fluid-solid systems, *Math. Comput.*, 67(1998), 221, 111-136.


B. Boroomand and O. C. Zienkiewicz, Recovery by equilibrium in patches (REP), *Int. J. Numer. Meth. Engrg.*, 40(1997), 137-164.


B. Boroomand and O. C. Zienkiewicz, An improved REP recovery and the effectivity robustness test, *Int. J. Numer. Meth. Engrg.*, 40(1997), 3247-3277.


S. W. Brady and A. R. Elcrat. Some results on a posteriori error estimation for approximate solution of second order elliptic problems, *SIAM J. Numer. Anal.*, 16(1979), 6, 877-889.


D. Braess. *Finite elements*. Cambridge University Press, Cambridge, 1997.


J.H. Bramble, R.D. Lazarov and J. E. Pasciak. Least-squares for second-order elliptic problems. *Comput. Methods Appl. Mech. Engrg.* , 152(1998), 195-210.


J. H. Bramble and R. S. Falk. A mixed-Lagrange multiplier finite element method for the polyharmonic equation, *RAIRO Model. Math. Anal. Numeric.* 19(4), 1985, 519-557.


H. Brezis. Problémes unilatéraux, *J. Math. Pures Appl.*, 9(1971, 1, 1-168.

- Brezis H., Kinderlehrer D. The smoothness of solutions to nonlinear variational inequalities, *Indiana Univ. Math J.*, 23(1974), 831-844.
- H. Brezis and M. Sibony. Equivalence de deux inéquations variationnelles et applications, *Arch. Rat. Mech. Anal.*, 41(1971), 254-265.
- F. Brezzi, On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers, *R.A.I.R.O., Annal. Numer.*, 8 (1974), R-2, 129-151.
- F. Brezzi, J. J. Douglas, R. Duran and M. Fortin. Mixed finite elements for second order elliptic problems in three variables, *Numer. Math.*, 51(1987), 2, 237-250.
- F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*, Springer Series in Computational Mathematics, 15, New York, 1991.
- F. Brezzi, M. Fortin and R. Stenberg. Error analysis of mixed-interpolated elements for Reissner-Midlin plates, *Math. Models and Methods in Applied Sciences*, 1(1991), 125-151.

F. Brezzi, C. Johnson, B. Mercier. Analysis of a mixed finite element method for elasto-plastic plates, *Mathematics of Computation,* (31)1977, 140, 809-817. Brezzi, F.; Pitkaranta, J. On the stabilization of finite element approximations of the Stokes equations. Efficient solutions of elliptic systems (Kiel, 1984), 11–19, *Notes Numer. Fluid Mech.*, 10, Vieweg, Braunschweig, 1984.

J. Bonvin, M. Picasso, and R. Stenberg, *GLS and EVSS methods for a three–field Stokes problem arising from viscoelastic flows*, Comput. Methods Appl. Mech. Engrg. 190(2001), 3893–3914.

I. G. Bubnov. *Selected Works.* Sudpromgiz, Leningrad (1956) (in Russian).

H. Buss and S. Repin. A posteriori error estimates for boundary-value problems with obstacles. In *Proceedings of 3d European Conference on Numerical Mathematics and Advanced Applications, Jyuvaskyla, 1999*, 162-170, World Scientific, 2000.

G. F. Carey and D. L. Humphrey. Mesh refinement and iterative solution methods for finite element computations, *Int. J. Numer. Meth. Engrg.*, 17(1981), 1717-1734.

G. F. Carey and A. I. Pehlivanov. Local error estimation and adaptive remeshing scheme for least-squares mixed finite elements, *Comput. Methods Appl. Mech. Engrg.*, 150(1997), 125-131.

C. Carstensen. A posteriori estimate for the mixed finite element method, *Math. Comput.*, 66(1997), 218, 465-476.

C. Carstensen. Quasi-interpolation and a posteriori error analysis of finite element methods, *Mathematical Modelling in Numerical Analysis*, 33(1999), 6, 1187-1202.

Carstensen, C.; Bartels, S. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. I: Low order conforming, nonconforming, and mixed FEM, *Math. Comp.*, 71(2002), 239, 945-969.


C. Carstensen and S. A. Funken. Fully reliable localized error control in the FEM, *SIAM J. Sci. Comput.*, 21(2000), 4, 1465-1484.


C. Carstensen and S. A. Funken. Constants in Clément-interpolation error and residual based a posteriori error estimates in Finite Element Methods, *East–West J. Numer. Math.*, 8(2000), 3, 153-175.


C. Carstensen and R. Verfürth. Edge residuals dominate a posteriori error estimates for low order finite element methods, *SIAM J. Numer. Anal.*, 36 (1999), 5, 1571-1587.


A. Charbonneau, K. Dossou and R. Pierre. A residual-based a posteriori error estimator for Ciarlet–Raviart formulation of the first biharmonic problem, *Numer. Meth. for PDE's*, 13(1997), 93-111.


Z. Chen and R. H. Nochetto. Rezidual type a posteriori error estimates for elliptic obstacle problems, *Numer. Math.*, 84(2000), 527-548.


X. Cheng, W. Han and H. Huang. Analysis of some mixed elements for the Stokes problem, *Journal of Computational and Applied Mathematics*, 85(1997), 19-35.

M. Chipot. Regularity for the two obstacles problem. In *Free boundary problems, Vol. II (Pavia, 1979), Ist. Naz. Alta Mat. Francesco Severi, Rome, 1980*, 135-140.

S. S. Chow. Finite element error estimates for non-linear elliptic equations of monotone type, *Numer. Math.*, 54(1989), 373-393.

S.-S. Chow, G. F. Carey and R. D. Lazarov. Natural and post-processed superconvergence in semilinear problems, *Numer. Meth. PDE*, 7(1991), 245-259.

P. G. Ciarlet. *The finite element method for elliptic problems*. North Holland, New York, Oxford, 1978.

Ph. Clément. Approximations by finite element functions using local regularization, *RAIRO Anal. Numér.*, 9(1975), R-2, 77-84.

Funktionanalysis und numerische mathematik. Springer-Verlag, Berlin, 1964.

P. Coorevits, J.-P. Dumeau and J.-P. Pelle. Error estimator and adaptivity for three-dimensional finite element analysis. In *Advances in*

Adaptive Computational Methods in Mechanics, Ed. P. Ladevéze and J. T. Oden, 443-457, Elsevier, New York, 1998.


R. Courant. Variational methods for some problems of equilibrium and vibration, *Bulletin of AMS*, 49(1943), 1-23.


Computational Methods in the mechanics of Fracture, Ed. S.N. Atluri, North–Holland, New York, 1986.


P. Destuynder and B. Métivet. Explicit error bounds of a conforming finite element method, *Math. Comput.*, 68(1999), 1379-1396.


W. Dörfler and M. Rumpf. An adaptive strategy for elliptic problems including a posteriori controlled boundary approximation, *Math. Comput.*, 67(1998), 224, 1361-1362.


J. Douglas, Jr. T. Dupont and L. B. Wahlbin. The stability in L^q of the L^2-projection into finite element function spaces, *Numer. Math.*, 23(1975), 193-197.

- R. Duran, M. A. Muschietti and R. Rodriguez. On the asymptotic exactness of error estimators for linear triangle elements, *Numer. Math.*, 59(1991), 107-127.
- R. Duran and R. Rodriguez. On the asymptotic exactness of Bank–Weiser's estimator, *Numer. Math.*, 62(1992), 297-303.
- G. Duvant and J.-L. Lions. *Les inequations en mecanique et en physique*, Dunod, Paris, 1972.
- D. C. Drucker and W. Prager. Soil mechanics and plastic analysis or limit design. *Quart. Appl. Math.*, 10(1952), 157-165.
- I. Ekeland and R. Temam. *Convex analysis and variational problems*. North-Holland, Amsterdam, 1976.
- H.W. Engl and O. Scherzer. Convergence rates results for iterative methods for solving nonlinear ill-posed problems. In *Surveys on solution methods for inverse problems*, 7-34, Springer-Verlag, Vienna, 2000.
- K. Eriksson, D. Estep, P. Hansbo and C. Johnson. *Computational differential equations*. Cambridge University Press, Cambridge, 1996.


K. Eriksson and C. Johnson. An adaptive finite element method for linear elliptic problems, *Math. Comput.*, 50(1988), 361-383.


K. Eriksson and C. Johnson. Adaptive finite element methods for parabolic problems. I. A linear model problem, *SIAM J. Numer. Anal.*, 28 (1991), 1, 43-77.


R. E. Ewing. A posteriori error estimation, *Comput. Meth. Appl. Mech. Engrg.*, 82(1990),1-3, 59-72.


R. E. Ewing, R. D. Lazarov and J. Wang, Superconvergence of the velocity along the Gauss lines in mixed finite element methods, *SIAM J. Numer. Anal.*, 18(1991), 1015-1029.


R. S. Falk. Error estimates for the approximation of a class of variational inequalities, *Math. Comput.*, 28(1974), 963-971.


M. Feistauer. *Mathematical methods in fluid dynamics*. Longman, Harlow, 1993.


W. Fenchel. On the conjugate convex functions, *Canad. J. Math.* 1(1949), 73-77.


Fraejs B. de Veubeke. Displacement and equilibrium models in the finite element methods, in *Stress analysis* (O. C. Zienkiewicz and G. S. Holister, eds.), Wiley and Sons, New-York, 1965, 145-197.


Fraejs B. de Veubeke. A conforming finite element for plate bending, *Internat. J. Solids and Structures* 4(1968), 95-108.


G. E. Forsythe, M. A. Malcolm and C. B. Moler. *Computer methods for mathematical computations*, Prentice-Hall, Englewood Cliffs, New York, 1977.


A. Friedman. *Variational principles and free-boundary problems*. Wiley and Sons, New York, 1982.


D. A. Field. Laplacian smoothing and Delanay triangulations, *Commum. Appl. Numer. Methods*, 4(1988), 709-712.


L. P. Franca, J. Karam Filho, A. F. D. Loula, and R. Stenberg, *A convergence analysis of a stabilized method for the Stokes flow,* Mat. Apl. Comput. 10 (1991),1, pp. 19–26.


M. Frolov, P. Neittaanmäki and S. Repin. On the reliability, effectivity and robustness of a posteriori error estimation methods. Reports of the Department of Mathematical Information Technology of the University of Jyväskylä, No. B14/2002.


M. Frolov, P. Neittaanmäki and S. Repin. On computational properties of a posteriori error estimates based upon the method of duality error majorants (to appear in Proc. 5th European Conference on Numerical Mathathematics and aplications, Praha,2003).


M. Fuchs and G.A. Seregin. *Variational methods for problems from plasticity theory and for generalized Newtonian fluids.* Lect. Notes in Mathematics 1749, Springer-Verlag, Berlin (2000).


H. Gaevskii, H., K. Gröger, and K. Zacharias. *Nichtlineare Operator-gliechungen and Operatordifferentialgleiichungen.* Akademie-Verlag, Berlin, 1974.


B. G. Galerkin. Beams and plates. Series in some questions of elastic equilibrium of beams and plates. *Vestnik Ingenerov*, 19(1915), 897-908 (in Russian).


L. Gallimard, P. Ladev eze and J. P. Pelle. Error estimation and adaptivity in elastoplasticity, *Int. J. Numer. Meth. Engrg.*, 39(1996), 189-217.


E.G. Geisler, A.A. Tal and D.P. Garg. On a-posteriori error bounds for the solution of ordinary nonlinear differential equations. In *Computers and mathematics with applications*, 407-416, Pergamon, Oxford, 1976.


C. I. Goldstain. Variational crimes and L^∞ error estimates in the finite element method, *Math. Comput.*, 35(1980), 152, 1131-1157.


D. Gilbarg and N.S. Trudinger. *Elliptic partial differential equations of second order*. Springer-Verlag, Berlin, 1977.


V. Girault and P. A. Raviart. *"Finite element approximation of the Navier–Stokes equations"*. Springer-Verlag, Berlin, 1986.


R. Glowinski. *Numerical methods for nonlinear variational problems*. Springer-Verlag, New-York, 1982.

R. Glowinski and O. Pironneau. Numerical Methods for the First Biharmonic Equation and for the Two-Dimensional Stokes Problem, *SIAM Review*, (21)1999, 2, 167-212.

R. Glovinski, J.-L.Lions, R. Trémolierès. *Analyse numérique des inéquations variationnelles*. Dunod, Paris, 1976.

R. Glowinski, T.-W. Pan and J. Periaux. A fictitious domain method for Dirichlet problem and applications, *Comput. Methods Appl. Mech. Engrg.*, 111(1994), 283-303.

W. Han. Finite element analysis of a holonomic elasto-plastic problem, *Numer. Math.*, 60(1992), 493-508.

W. Han. A posteriori error analysis for linearization of nonlinear elliptic problems and their discretization, *Math. Meth. Appl. Sci.* , 17(1994), 487-508.

W. Han and D. C. Reddy. On the finite element method for mixed variational inequalities arising in elastoplasticity, *SIAM J. Numer. Anal.*, 32(1995), 6, 1776-1807.

- W. Han and D. Reddy. Qualitative and numerical analysis of quasi-static problems in elastoplasticity, *SIAM J. Numer. Anal.*, 34(1997), 1, 143-177.
- W. Han, S. Jensen and I. Shimansky. The Kačanov method for some nonlinear problems, *Applied Numerical Mathematics*, 24(1997), 57-79.
- P. Hansbo. Generalized Laplacian smoothing of unstructured grids, *Commun. Numer. Methods Eng.*, 11(1995), 455-464.
- Y. Hayashi. On a posteriori error estimation in the numerical solution of systems of ordinary differential equations. *Hiroshima Math. J* 9(1979), no, 1, 201-243.
- R.Hill. *The mathematical theory of plasticity.* Oxford, 1983
- I. Hlaváček and M. Křižek. On a superconvergence finite element scheme for elliptic systems. I. Dirichlet boundary conditions. *Aplikace Matematiky*, 32(1987), No.2, 131-154.
- R. H. W. Hoppe. Mortar finite elements in R^3, *East-West J. Numer. Math.*, 7(1999), 3, 159-173.


R. H. W. Hoppe and R. Kornhuber. Adaptive multilevel methods for obstacle problems. *SIAM J. Numer. Anal.*, 31(1994), 301-323.


P. Houston, R. Rannacher, E. Süli. A posteriori error analysis for stabilised finite element approximations of transport problems, *Comput. Methods Appl. Mech. Engrg.* 190, 1483-1508, 2000.


A. A. Iljushin. *Plasticite. (Deformations elastico-plastiques)*. (Eyrolles, Paris, 1956).


A.D. Ioffe and V.M. Tikhomirov. *Theory of extremal problems. Studies in Mathematics and its Applications, 6.* North-Holland, Amsterdam-New York, 1979


J.L. Jensen. Sur les fonctions convexes et les inegalités entre les valeurs moyennes, *Acta Math.*, 30(1906), 175-193.


Jin Qi-nian. Error estimates of some Newton-type methods for solving nonlinear inverse problems in Hilbert scales, *Inverse Problems*, 16(2000), 187-197


C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Methods*, Cambridge University Press, Cambridge, 1987.


C. Johnson. On Finite Element Methods for Plasticity Problems.- *Numer. Math.*, 26(1976), 79-84.


C. Johnson. Adaptive finite element methods for diffusion and convection problems, *Comput. Methods Appl. Mech. Engrg.*, 82(1990), 301-322.


C. Johnson and P. Hansbo. Adaptive finite elements in computational mechanics, *Comput. Methods Appl. Mech. Engrg.* 101(1992), 143-181.


C.Johnson and B.Mercier. Some equilibrium finite element methods two-dimensional elasticity problems, *Numer. Math.*, 30(1978), 101-116.


C. Johnson and R. Rannacher. On error control in computational fluid dynamics (CFD) Preprint 1994–07, Chalmers University of Technology, Göteborg.


C. Johnson, R. Rannacher and M. Boman. Numerics in hydrodynamic stability. Towards error control in CFD, *SIAM J. Numer. Anal.*, 32(1995), 1058-1079.


C. Johnson, R. Rannacher and M. Boman. *On transition to turbulence and error control in CFD*, Preprint 1994–26, Chalmers University of Technology, Göteborg.


C. Johnson and A. Szepessy. Adaptive finite element methods for conservation laws based on a posteriori error estimates, *Commun. Pure and Appl. Math.*, vol. XLVIII (1995), 199-234.


B.-O. Heimsund, X.-C. Tai and J. Wang. Superconvergence for the gradient of finite element approximations by L^2 projections, *SIAM J. Numer. Anal.*, 40(2002), 4, 1263-1280.


H. Kavarada. *Numerical problems for free surface problems by means of penalty*. Lecture Notes in Mathematics, No 704, Springer-Verlag, 1979.


H. Kardestuncer and D. H. Norrie ed. *Finite Element Handbook*. McGraw–Hill, New York, 1987.


D. W. Kelly. The self equilibration of residuals and complementary error estimates in the finite element method, *Internat. J. Numer. Meth. Engrg.* 20(1984) 1491-1506.


D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*. Academic Press, New York, 1980.


D. W. Kelly, J. R. Gago, O. C. Zienkiewicz and I. Babuška. A posteriori error analysis and adaptive processes in the finite element method. Part I - error analysis, *Internat. J. Numer. Meth. Engrg.*, 19(1983), 1593-1619.


R. V. Kohn. The relaxation of a double-well energy, *Continuum Mech. Thermodynamics*, 3(1991), 3, 193-236.


A. N. Kolmogorov and S. V. Fomin. *Introductory real analysis*. Dover Publications, Inc., New York, 1975.


R. Kornhuber. A posteriori error estimates for elliptic variational inequalities, *Comput. Math. Appl.*, 31(1996), 49-60.


S. Korotov, P. Neittaanmaki and S. Repin. A posteriori error estimation of goal-oriented quantities by the superconvergence patch recovery, *J. Numer. Math.* 11 (2003), 1, 33-59.


M. Křížek, P. Neittaanmäki and R Stenberg eds. *Finite Element Methods. Superconvergence, Post-Processing and A Posteriori Error Estimates*. Lecture Notes in Pure and Applied Mathematics, Vol. 196, Marcel Dekker, New York, 1998.


M. Křížek and P. Neittaanmäki. Superconvergence phenomenon in the finite element method arising from averaging of gradients *Numer. Math.*, 45(1984), 105-116.


P. Ladeveze and D. Leguillon. Error estimate procedure in the finite element method and applications, *SIAM J. Numer. Anal.*, 20(1983), 485-509.


P. Ladeveze, J.-P. Pelle and Ph. Rougeot. Error estimation and mesh optimization for classical finite elements, *Engineering Computations*, 8(1991), 69-80.


P. Ladev eze and Ph. Rougeot. New advances on a posteriori error on constitutive relation in f.e. analysis, *Comput. Methods Appl. Mech. Engrg.*, 150(1997), 239-249.


O. A. Ladyzhenskaya. *Mathematical problems in the dynamics of a viscous incompressible fluid*. Nauka, Moscow, 1970 (in Russian).


O. A. Ladyzhenskaya, *The boundary value problems of mathematical physics*. Springer-Verlag, New York, 1985.


O. A. Ladyzhenskaya, On mosified Navier–Stokes equations for large velocity gradients, *Zapiski Nauchnych Seminarov LOMI*, 7(1968), 126-154.


O. A. Lady zenskaja and V. A.; Solonnikov. Some problems of vector analysis, and generalized formulations of boundary value problems for the Navier-Stokes equation, *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 59(1976), 81–116, 256 (Russian).


O. A. Ladyzhenskaya and N. N. Uraltseva. *Linear and Quasilinear Elliptic equations*, Academic Press, New York, 1968.

R. D. Lazarov. Superconvergence of the gradient for triangular and tetrahedral finite elements of a solution of linear problems in elasticity theory, *Computational Processes and Systems*, 6(1988), 180-191 (in Russian).

A.S. Leonov. Some a posteriori stopping rules for iterative methods for solving linear ill-posed problems, *Comput. Math. Math. Phys.*, 34(1994), 1, 121-126.

J.-L. Lions and E. Magenes. Problḿes aux limites non homogénes et applications. Dunod, Paris, 1968.

R. Löhner, K. Morgan and O. C. Zienkiewicz. Adaptive grid refinement for the compressible Euler equations. In *Accuracy Estimates and Adaptive Refinements in Finite Element Computations, I. Babuška, O. C. Zienkiewicz, J. Gago and E. R. de Olivieira eds., Wiley and Sons, 1986*, 281-297.

J. Malek, J. Nečas, J. Rokuta and M. Ružička. *Weak and measure valued solutions to evolution partial differential equations. Applied*

Mathematic and Mathematical Computation vol 13., Chapman and Hall, 1996.


J. Medina, M. Picasso and J. Rappaz. Error estimates and adaptive finite elements for nonlinear diffusion-convection problems. Ecole Polytechnique Federale de Lausanne, Preprint CH-1015, 1995.


P. Meyer. A unifying theorem on Newton's method, *Numer. Funct. Anal. Optim.* 13(1992), no. 5-6, 463-473.


S. G. Mikhlin. *Variational methods in mathematical physics*. Pergamon, Oxford, 1964.


S. G. Mikhlin. *Error Analysis in Numerical Processes* Wiley and Sons, Chiester–New York, 1991.


S. G. Mikhlin. *Constants in some inequalities of analysis*. Wiley and Sons, Chiester–New York, 1986.


P.Mosolov and V.Myasnikov. *Mechanics of rigid plastic bodies*, Nauka, Moscow, 1981 (in Russian).


A. Muzalevsky and S. Repin. On two-sided error estimates for approximate solutions of problems in the linear theory of elasticity, *Russian J. Numer. Anal. Math. Modelling*, 18(2003), 1, 65-85.


P. Neittaanmäki and M. Křížek. On $0(h^4)$ superconvergence of piecewise bilinear FE-approximations. In Teubner-Texte Math. 107, Teubner, 1988, 250-255.


P. Neittaanmäki and S. Repin. A posteriori error estimates for boundary-value problems related to the biharmonic operator, *East-West J.Numer. Math.*, 9(2001), 2, 157-178.


P. Neittaanmäki and S. Repin, *Reliable methods for computer simulation, Error control and a posteriori estimates*, Elsevier, New York, 2004.


J. T. Oden, L. Demkowicz, T. Strouboulis and P. Devloo. Adaptive methods for problems in solid and fluid mechanics. In *Accuracy Estimates and Adaptive Refinements in Finite Element Computations, I. Babuška, O. C. Zienkiewicz, J. Gago and E. R. de Oliveira eds.*, Wiley and Sons, 1996.


J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Westermann. Towards a universal $h - p$ adaptive finite element strategy. Part 2. A posteriori error estimation, *Comput. Methods Appl. Mech. Engrg.*, 77 (1989) 113-180.


J. T. Oden, S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element method, *Comput. Math. Appl.*, 41, 735-756, 2001.


J. T. Oden and J. N. Reddy. *Mathematical theory of finite elements*. Wiley and Sons, 1976.


J. T. Oden, W. Wu and M. Ainsworth. An a posteriori error estimate for finite element approximations of the Navier–Stokes equations, *Comput. Methods. Appl. Mech. Engrg.*, 111(1994), 185-202.


L. A. Oganesjan and L. A. Ruchovec. An investigation of the rate of convergence of variation-difference schemes for second order elliptic equations in a two-dimensional region with smooth boundary, *Z. Vychisl. Mat. i Mat. Fiz.*, 9(1969), 1102–1120, (Russian).

A. Ostrowski. Les estimations des erreurs a posteriori dans les procédés itératifs, *C.R. Acad.Sci. Paris Sér. A–B*, 275(1972), A275-A278.

D.R.J.Owen and E.Hinton. *Finite Elements in Plasticity Theory: Theory and Practice*. Prinerdge Press, Swansea, 1980.

C. Padra, A posteriori error estimators for nonconforming approximation of some quasi-newtonian flows, *SIAM J. Numer. Anal.*, 34(1997), 1600-1615.

J. Peraire and A. T. Patera. Bounds for linear-functional outputs of coercive partial differential equations: Local indicators and adaptive refinement. In *Advances in Adaptive Computational Methods in Mechanics, Ed. P. Ladevéze and J. T. Oden, Elsevier, New York, 1998*, 199-228.

F. Potra. Sharp error bounds for a class of Newton-like methods, *Libertas Math.* 5(1985), 71-84

J. Pousin and J. Rappaz. Consistance, stabilité, erreurs a priori et a posteriori pour des probl emes non linéaries. *C. R. Acad. Sci. Paris* 312(1991), 699-703.


J. Pousin and J. Rappaz. Consistency, stability, a priori and a posteriori errors for Petrov-Galerkin methods applied to nonlinear problems. *Numer. Math.* 69(1994), 213-231.


W. Prager and J. L. Synge. Approximation in elasticity based on the concept of function space, *Quart. Appl. Math.* 5(1947), 241-269.


R. Rannacher. Zur L^{∞}-Konvergenz linear finite elemente beim Dirichlet Problem, *Math. Z.*, 149(1976), 69-77.


R. Rannacher. On nonconforming and mixed finite element method for plate bending problems, the linear case, *R.A.I.R.O. Anal. Numer.*, 13(1979), 4, 369-387.


R. Rannacher and R. Scott. Some optimal error estimates for piecewise linear finite element approximations, *Math. Comput.*, 38(1982), 158, 437-445.


R. Rannacher and F. T. Suttmeier. A feed–back approach to error control in finite element methods: application to linear elasticity. IWR, Preprint 96–42(SFB 359), Heidelberg 1996.


R. Rannacher and F.T. Suttmeier. A posteriori error control and mesh adaptation for F.E. models in elasticity and elasto-plasticity. In *Advances in Adaptive Computational Methods in Mechanics, Eds. P. Ladevéze and J. T. Oden, Elsevier, New York, 1998*, 275-292.


W. C. Rheinboldt. On a theory of mesh-refinement processes, *SIAM J. Numer. Anal.*, 17(1980), 766-778.


S. Repin. A posteriori estimates for approximate solutions of variational problems with strongly convex functionals, *Problems of Mathematical Analysis*, 17 (1997), 199-226. (in Russian, translated in *Journal of Mathematical Sciences*, 97(1999), 4, 4311-4328).


S. Repin. A posteriori estimates of the accuracy of variational methods for problems with nonconvex functionals, *Algebra i Analiz*, 11(1999), 4, 151-182 (in Russian, translated in *St.-Petersburg Mathematical Journal*, 11(2000), 4, 651-672).


S. Repin. A posteriori error estimation for nonlinear variational problems by duality theory. *Zapiski Nauchnych Seminarov POMI*, 243(1997) 201-214.


S. Repin. A posteriori error estimation for variational problems with power growth functionals based on duality theory, *Zapiski Nauchnych Seminarov POMI*, 249(1997), 244-255.


S. Repin. A posteriori error estimates for approximate solutions of variational problems. In *Proceedings of 2nd European Conference on Numerical Mathematics and Advanced Applications, Heidelberg, 1997*, 524-531, World Sci. Publishing, River Edge, New York, 1998.


S. Repin. A unified approach to a posteriori error estimation based on duality error majorants, *Mathematics and Computers in Simulation*, 50(1999), 313-329.


S. Repin. A posteriori error estimation for variational problems with uniformly convex functionals, *Math. Comput.*, 69(230), 2000, 481-500.


S. Repin. Two-sided estimates for deviation from an exact solution to uniformly elliptic equation. *Trudi St.-Petersburg Math. Society*, 9(2001), 148-179 (in Russian, translated in *American Mathematical Translations Series 2, 9(2003)*)


S.Repin. Estimates of deviation from exact solutions of initial-boundary value problems for the heat equation, *Rend. Mat. Acc. Lincei*, 13(2002), 121-133.


S. Repin. Estimates of deviations from exact solutions of elliptic variational inequalities, *Zapiski Nauchn. Semin. V.A. Steklov Mathematical Institute in St.-Petersburg (POMI)*, 271(2000), 188-203.


S. Repin. Aposteriori estimates for the Stokes problem, *Journal of Math. Sciences* 109 (2002), 5, 1950-1964.


S. Repin. Estimates of deviations for generalized Newtonian fluids, *Zapiski Nauchn. Semin. V.A. Steklov Mathematical Institute in St.-Petersburg (POMI)*, 288(2002), 178-203.


S. Repin. Estimates for errors in two-dimensional models of elasticity theory, *J. Math. Sci. (New York)*, 106 (2001), no. 3, 3027-3041.


S. Repin, *Functional Approach to Locally Based A Posteriori Error Estimates for Elliptic and Parabolic Problems*, Proc. 6th European

Conference on Numerical Mathematics and Advanced Applications (Santiago de Compostela) 2005, Springer, Berlin, 2006, pp. 133-148.

S. Repin, *Local a posteriori estimates for the Stokes problem,* Zap. Nauchn, Sem. S.-Peterburg, Otdel. Mat. Inst. Steklov. (POMI) 318 (2004), pp. 233–245.

S. Repin and M. Frolov. A posteriori error estimates for approximate solutions of elliptic boundary value problems, *Zh. Vychisl. Mat. Mat. Fiz.*, 42(2002), 12, 1774–1787 (Russian).

S. Repin, S. Sauter and A. Smolianski. A posteriori error estimation for the Dirichlet problem with account of the error in the approximation of boundary conditions, *Computing*, 70(2003), 205-233.

S. Repin, S. Sauter and A. Smolianski. Duality Based A Posteriori Error estimator for the Dirichlet Problem, *Proc.Appl. Math. Mech.*, 2 (2003), 513-514.

S. Repin, S. Sauter and A. Smolianski. A posteriori error estimation of dimension reduction errors (to appear in Proc. 5th European

Conference on Numerical Mathathematics and aplications, Praha,2003).


S. Repin, S. Sauter and A. Smolianski, *A posteriori estimation of dimension reduction errors for elliptic problems in thin domains*, *SIAM J. Numer. Anal.*, 42 (2004), 4, pp. 1435–1451.


S. Repin, S. Sauter and A. Smolianski, *A Posteriori Control of Dimension Reduction Errors on Long Domains*, Proceedings in Applied Mathematics and Mechanics, 4, 1, 714–715 (2004).


S. I. Repin and L. S. Xanthis. A posteriori error estimation for elasto-plastic problems based on duality theory, *Comput. Methods Appl. Mech. Engrg.*, 138(1996), 317-339.


S. I. Repin and L. S. Xanthis. A posteriori error estimation for nonlinear variational problems, *Comptes Rendus de l'Académie des Sciences, Mathématique*, 324(1997), 1169-1174.


W. Ritz. Über eine neue Methode zur Lözing gewisser Variationsprobleme der Mathematischen Physics, *J. Reine Angew. Math.*, 135(1909), 1-61.


R. T. Rockafellar. *Convex analysis*, Princeton Univ. Press, 1970.


R. Rodrigues. Some remarks on Zienkiewicz–Zhu estimator, *Numer. Methods for PDE*, 10(1994), 625-635.


A. H. Schatz. Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids: Part 1: global estimates, *Math. Comput.*, 67(1998), 223, 877-899.


A. H. Schatz and L. B. Wahlbin. Interior maximum norm estimates for finite element methods, *Math. Comput.*, 31(1977), 414-442.


A. H. Schatz and L. B. Wahlbin. Maximum norm estimates for finite element method on plane polygonal domains. Part 1, *Math. Comput.*, 32(1978), 73-109.


A. H. Schatz and L. B. Wahlbin. Maximum norm estimates for finite element method on plane polygonal domains. Part 2, refinements, *Math. Comput.*, 33(1979), 465-492.


C. Schwab. A-posteriori modelling error estimation for hierarchic plate models, *Numer. Math.*, 74(1996), 221-259.


A. H. Schatz and J. Wang. Some new error estimates for Ritz–Galerkin methods with minimal regularity assumptions, *Math. Comput.*, 65(1996), 213, 19-27.


M. Schultz and O.Steinbach. A new a posteriori error estimator in adaptive direct boundary element methods: the Dirichlet problem, *Calcolo*, 37(2000), 79-96.


M. Schulz and W. L. Wendland. A general approach to a posteriori error estimates for strictly monotone and Lipschitz continuous nonlinear operators illustrated in elasto-plasticity. In *Proceedings of 2nd European Conference on Numerical Mathematics and Advanced Applications, Heidelberg, 1997, World Sci. Publishing, River Edge, New York, 1998*, 572-579.


M. Schultz and W. Wendland. *On an adaptive finite element method for elasto-plastic deformation with hardening.* Preprint 1997/39, SFB 259, Stuttgardt, 1997.


S. L. Sobolev. *Some Applications of Functional Analysis in Mathematical Physics*, Izdt. Leningrad. Gos. Univ., Leningrad, 1955 (in Russian

translated in *Translation of Mathematical Monographs, Volume 90* American Mathematical Society, Providence, RI, 1991).

E. Stein and S. Ohnimus. Coupled model- and solution-adaptivity in the finite element method, *Comput. Methods Appl. Mech. Engrg.*, 150(1997), 327-350.

E. Stein, F.J. Bartold, S. Ohnimus and M. Schmidt. Adaptive finite elements in elastoplasticity with mechanical error indicators and Neumann-type estimators. In *Advances in Adaptive Computational Methods in Mechanics, Ed. P. Ladevéze and J. T. Oden, Elsevier, New York, 1998*, 81-100.

G. Strang and G. Fix. *An analysis of the finite element method.* Prentice Hall, Englewood Cliffs, 1973.

T. Strouboulis and J. T. Oden. A posteriori error estimation of finite element approximations in fluid mechanics, *Comput. Meth. Appl. Mech. Engrg.*, 78(1990), 201-242.

- J. L. Synge. The hypercircle method. In *Studies in numerical analysis (papers in honour of Cornelius Lanczos on the occasion of his 80th birthday)*, 201-217. Academic Press, London, 1974.
- R. Temam, *Navier–Stokes equations. Theory and numerical analysis,* Studies in Mathematics and its Applications, 2, North-Holland, Amsterdam, 1979.
- K. Tsuruta and K. Ohmori. A posteriori error estimation for Volterra integro-differential equations, *Mem. Numer. Math. No. 3* 1976, 33-47.
- N.N. Uraltseva. On the regularity of solutions to variational inequalities *Uspekhi Mat. Nauk*, 42(1987), 6, 151-174 (in Russian).
- R.S. Varga. *Matrix Iterative Analysis.* Prentice–Hall, Englewood Cliffs, New Jersey, 1962.
- A. Veeser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems, *SIAM J. Numer. Anal.*, 39(2001), 146-167.
- R. Verfürth. A posteriori error estimators for the Stokes equations, *Numer. Math.*, 55(1989), 309-326.


R. Verfürth. A posteriori error estimates for nonlinear problems. Finite element discretisations of elliptic equations, *Math. Comput.* 62(1994) 445-475.


R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques* Wiley and Sons, Teubner, New-York, 1996.


R. Verfürth. A posteriori error estimates for nonlinear problems. $L^r(0,T;L^p(\Omega))$-error estimates for finite element discretizations of parabolic equations, *Math. Comput.*, 67(1998), 224, 1335-1360.


A. Vesser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems, *SIAM J. Numer. Anal.*, 39(2001), 1, 146-167.


M. I. Vishik . The method of orthogonal projections for selfadjoint equations , *Sov. Math. Dokl* , 1947, 56.


L. B. Wahlbin. *Superconvergence in Galerkin Finite Element Methods*, Lecture Notes in Mathematics, No 1605, Springer-Verlag, 1995.


J. Wang. Superconvergence analysis of finite element solutions by the least-squares surface fitting on irregular meshes for smooth problems, *J. Math. Study*, 33(2000), 3, 229-243.


H. Weil. The method of orthogonal projections in potential theory, *Duke Math. J.*, 7(1940), 411-444.


J. R. Whiteman and G. Goodsell. A survey of gradient superconvergence for finite approximations to second order elliptic problems on triangular and tetrahedral meshes. In *The Mathematics of Finite Elements and Applications VII, Ed. J. R. Whiteman, Academic Press, 1991*, 55-74.


W. Wunderlich, H. Gramer and G. Steinl. An adaptive finite element approach in associated and non-associated plasticity considering localization phenomena. In *Advances in Adaptive Computational Methods in Mechanics*, Ed. P. Ladevéze and J. T. Oden, 293-308, Elsevier, New York, 1998.


On classes of summable functions and their Foutier series, *Proc. Roy. Soc. Ser. A*, 87(1912), 225-229.

S. Zaremba. Sur un probleme toujours possible comprenant, a titre de cas particuliers, le probleme de Dirichlet et celui de Neumann, *J. math. pures et appl.* 6(1927),2, 127-163.

E. Zeidler. *Nonlinear functional analysis and its applications. I. Fixed-point theorems.* Springer-Verlag, New York, 1986,

O. C. Zienkiewicz. Achievements and some unsolved problems of the finite element method, *Int. J. Numer. Meth. Engrg.*, 47(2000), 9-28.

O. C. Zienkiewicz, B. Boroomand and J.Z. Zhu. Recovery procedures in error estimation and adaptivity: Adaptivity in linear problems. In *Advances in Adaptive Computational Methods in Mechanics, Ed. P. Ladevéze and J. T. Oden, Elsevier, New York, 1998*, 3-23.

O. C. Zienkiewicz and K. Morgan. *Finite elements and approximation*. Wiley and Sons, New York, 1983.

O.C.Zienkiewicz, S.Valliappan, I.P.King. Elasto-plastic solutions of engineering problems, initial stress finite element approach.- *Int. J. Numerical Methods Engrg.*, 1969, 1, 75-100.


O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis, *Internat. J. Numer. Meth. Engrg.*, 24(1987) 337-357.


O. C. Zienkiewicz and J. Z. Zhu. Adaptive techniques in the finite element method, *Commun. Appl. Numer. Methods*, 4(1988), 197-204.


O. C. Zienkiewicz and J. Z. Zhu. The superconvergent patch recovery and a posteriori error estimates. Part 1: The recovery technique, *Int. J. Numer. Meth. Engrg.*, 33(1992), 1331-1364.


J. Z. Zhu. A posteriori error estimation - the relationship between different procedures, *Comput. Methods Appl. Mech. Engrg.*, 150(1997), 411-422.


M. Zlámal. Some superconvergence results in the finite element method. Mathematical aspects of finite element methods. In *Proc. Conf., Consiglio Naz. delle Ricerche (C.N.R.), Rome, 1975. Lecture Notes in Math., Vol. 606, Springer-Verlag, Berlin, 1977*, 353-362.