LECTURES ON A POSTERIORI ERROR CONTROL

History. Error indicators for finite element methods.

Functional a posteriori error estimates.

A posteriori estimates for the Stokes problem.

A posteriori estimates for the linear elasticity problem.

A posteriori estimates for mixed methods.

Evaluation of errors arising due to data indeterminacy.

A posteriori estimates for iteration methods.

Functional a posteriori estimates for variational inequalities.

S. Repin,

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

Linz, December 2005

1. Lecture 1. INTRODUCTION. ERROR ANALYSIS IN THE MATHEMATICAL MODELING
   - Errors in mathematical modeling
   - Mathematical background
   - A priori estimates
2. Lecture 2. A CONCISE OVERVIEW OF A POSTERIORI ERROR ESTIMATION METHODS FOR APPROXIMATIONS OF DIFFERENTIAL EQUATIONS.
   - First approaches
   - Residual method
   - A posteriori methods based on post–processing
   - A posteriori methods using adjoint problems
3. Lecture 3. FUNCTIONAL A POSTERIORI ESTIMATES. FIRST EXAMPLES.
   - Functional A Posteriori Estimates. The concept
   - Derivation by the variational method
   - Derivation by the method of integral identities
   - Properties of functional a posteriori estimates
   - How to use them in practice?

7 Lecture 7. FUNCTIONAL A POSTERIORI ESTIMATES. STOKES PROBLEM.

- Stokes problem
- Inf-Sup condition
- Existence of a saddle point
- Estimates of the distance to the set of solenoidal fields
- Comments on the value of the LBB–constant
- Functional a posteriori estimates for the Stokes problem
- Estimates for problems with condition $\mathbf{div} u = \phi$.
- Problems for almost incompressible fluids
- Generalizations to problems where a solution is seeking in a subspace

8 Lecture 8. ESTIMATION OF INDETERMINACY ERRORS.

- Errors arising due to data indeterminacy. Examples
- General concept
- Upper bound of the error
- Lower bound of the error

9 Lecture 9. A POSTERIORI ESTIMATES FOR MIXED METHODS.

- Mixed approximations. A glance from the minimax theory
- A priori error analysis for the dual mixed method
- A posteriori error estimates for the primal mixed method

- A posteriori error estimates for the dual mixed method

10 Lecture 10. A POSTERIORI ERROR ESTIMATES FOR ITERATION METHODS

- Banach fixed point theorem
- A posteriori estimates
- Advanced two–sided a posteriori estimates
- Applications to problems with bounded linear operators
- Iteration methods in linear algebra
- Positivity methods and a posteriori error bounds.
- Applications to integral equations
- Applications to ordinary differential equations

11 Lecture 11. A POSTERIORI ESTIMATES FOR VARIATIONAL INEQUALITIES

- Variational inequalities
- Obstacle problem. Introduction
- Deviation estimate for variational inequalities
- Problems with two obstacles
- Perturbed problem
- A posteriori estimates for problems with two obstacles
- Numerical examples

- The elasto-plastic torsion problem

12 Lecture 12. FUNCTIONAL A POSTERIORI ESTIMATES FOR NONLINEAR VARIATIONAL PROBLEMS

- General form of the functional a posteriori estimate
- Problems with linear functional
- Examples
- Evaluation of errors in terms of local quantities
- Error estimation of modeling errors

Preface

This lecture course was prepared for the Special Radon Semester organized in October–December 2005 by J. Radon Institute of Computational and Applied Mathematics (RICAM) in Linz, Austria.
The main purpose of the course is to present (at least for certain classes of partial differential equations) a mathematically justified and practically efficient answer to the question:

How to verify the accuracy of approximate solutions computed by various numerical methods ?

During the last decade, this question has been intensively investigated by the functional methods of the theory of partial differential equations. As a result a new (functional) approach to the a posteriori error control of differential equations has been formed. In the present course of lectures, I tried to present the main ideas and results of this approach in the most transparent form and discuss it using several classical problems (diffusion problem, linear elasticity, Stokes problem) as basic examples.

The material is based on earlier lectures on a posteriori estimates and adaptive methods (University of Houston (2002), USA; Summer Schools of the University of Jyväskylä, Finland (2003, 2005); St.-Petersburg Polytechnical University). Also, I used some publications appeared in 2000-2004. However, in many parts the course is quite new and reflects the latest achievements in the area. A list of the literature 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 am grateful to RICAM and especially to Prof. U. Langer for the kind support. Also, I thank Prof. D. Braess, Prof. R. Lazarov, Dr. J. Valdman, and Dr. S. Tomar for the interest and discussions.

Sergey Repin Linz, December 2005

Lecture 1.
INTRODUCTION. ERROR ANALYSIS IN THE MATHEMATICAL MODELING

Lecture plan

- Errors arising in mathematical modeling;
- Basic mathematical knowledge
  - Notation
  - Functional spaces and inequalities;
  - Generalized solutions.
- A priori error estimates for elliptic type PDE's

We begin with two assertions that present a motivation of this lecture course.

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, but to compute their approximate solutions and analyze them.

II. Approximate solutions contain errors of various nature.

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

III. Error analysis of the approximate solutions to differential equations is one of the key questions in the Mathematical Modeling.

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

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

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

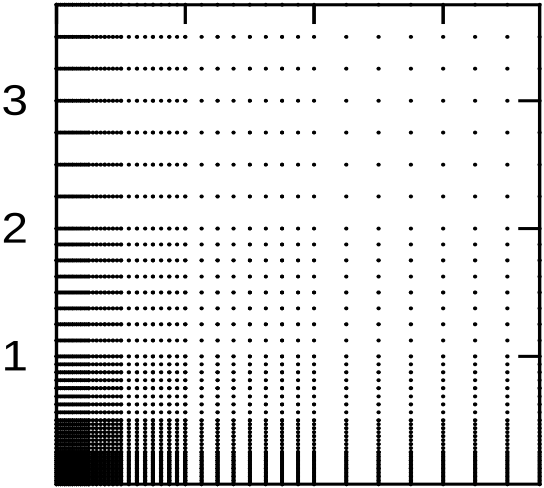
Numbers in a computer are presented in a floating point format:

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

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

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

The set $\mathbf{R}_{\mathbf{q}\ell\mathbf{k}} \times \mathbf{R}_{\mathbf{q}\ell\mathbf{k}}$



Example

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

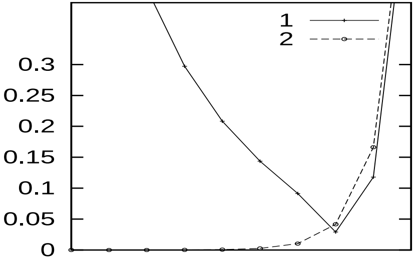
$$\mathbf{b} = \left(0 + \frac{1}{2} + 0\right) * 2^2 = \left(0 + 0 + \frac{1}{2}\right) * 2^3 = (0 + 0 + 0) * 2^4$$

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

Definition. The smallest floating point number which being added to 1 gives q quantity different from 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

$$
\begin{array}{llll}
\mathbf{U} & \textbf{Physical object/process} & & \\
 & \Downarrow & & \\
 & \varepsilon_1 & \longrightarrow & \boxed{\textbf{Error of a model}} \\
 & \Downarrow & & \\
\mathbf{u} & \textbf{Differential model} & & \mathbf{A}\mathbf{u} = \mathbf{f} \\
 & \Downarrow & & \\
 & \varepsilon_2 & \longrightarrow & \boxed{\textbf{Approximation error}} \\
 & \Downarrow & & \\
\mathbf{u_h} & \textbf{Discrete model} & & \mathbf{A^h u_h} = \mathbf{f_h} \\
 & \Downarrow & & \\
 & \varepsilon_3 & \longrightarrow & \boxed{\textbf{Computational error}} \\
 & \Downarrow & & \\
\mathbf{u_h^\varepsilon} & \textbf{Numerical solution} & & \mathbf{A^h u_h^\varepsilon} = \mathbf{f_h} + \epsilon.
\end{array}
$$

Two principal relations

I. Computations on the basis of a reliable (certified) model. Here ε_1 is assumed to be small and $\mathbf{u}_h^\varepsilon$ gives a desired information on $\mathbf{U}$.

$$\|\mathbf{U} - \mathbf{u}_h^\varepsilon\| \leq \quad \boldsymbol{\varepsilon_1} + \boxed{\varepsilon_2 + \varepsilon_3}. \tag{1.1}$$

II. Verification of a mathematical model. Here physical data $\mathbf{U}$ and numerical data $\mathbf{u}_h^\varepsilon$ are compared to judge on the quality of a mathematical model

$$\|\boldsymbol{\varepsilon_1}\| \leq \|\mathbf{U} - \mathbf{u}_h^\varepsilon\| + \boxed{\varepsilon_2 + \varepsilon_3}. \tag{1.2}$$

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.

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

Does such a function $\mathbf{u}$ exists and unique? 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

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

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

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

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

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

Unit matrix is denoted by $\mathbb{I}$. If $\boldsymbol{\tau} \in \mathbb{M}^{\mathbf{n}\times\mathbf{n}}$, 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

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

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
$\mathbf{C^k}(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}})$ $(\mathbf{C^k}(\boldsymbol{\Omega}, \mathbb{M}^{\mathbf{n}\times\mathbf{n}}))$,
$\mathbf{L^p}(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}})$ $(\mathbf{L^p}(\boldsymbol{\Omega}, \mathbb{M}^{\mathbf{n}\times\mathbf{n}}))$
with similar norms.

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

Generalized derivatives

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

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

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

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

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 $\mathbf{D}^{\alpha}\mathbf{f}$ in place of $\partial^{\mathbf{k}}\mathbf{f}/\partial\mathbf{x}_{\mathbf{1}}^{\alpha_1}\partial\mathbf{x}_{\mathbf{2}}^{\alpha_2}\ldots\partial\mathbf{x}_{\mathbf{n}}^{\alpha_\mathbf{n}}$.

Definition

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

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

Then, **g** is called a **generalized derivative** of **f** of degree

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

and we write

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

Sobolev spaces

The spaces of functions that have integrable generalized derivatives up to a certain order are called Sobolev spaces.

Definition

$\mathbf{f} \in \mathbf{W}^{1,p}(\Omega)$ if $\mathbf{f} \in \mathbf{L}^{p}$ and all the generalized derivatives of $\mathbf{f}$ of the first order are integrable with power 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:

$$\|\mathbf{f}\|_{1,p,\Omega} := \left(\int\limits_{\Omega} (|\mathbf{f}|^{p} + \sum_{i=1}^{n} |\mathbf{f}_{,i}|^{p}) d\mathbf{x} \right)^{1/p}.$$

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

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

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

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

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

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

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

where **c** is a positive constant independent of **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 $\mathbf{L}^2(\partial\boldsymbol{\Omega})$, which is the space $\mathbf{H}^{1/2}(\partial\boldsymbol{\Omega})$. 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

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

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

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

Inequalities

In the lectures we will use the following inequalities **1. Friederichs-Steklov inequality.**

$$\|\mathbf{w}\| \le \mathbf{C}_{\boldsymbol{\Omega}} \|\nabla \mathbf{w}\|, \quad \forall \mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^{1}(\boldsymbol{\Omega}), \tag{1.3}$$

2. Poincaré inequality.

$$\|\mathbf{w}\| \le \widetilde{\mathbf{C}}_{\boldsymbol{\Omega}} \|\nabla \mathbf{w}\|, \quad \forall \mathbf{w} \in \widetilde{\mathbf{H}}^{1}(\boldsymbol{\Omega}), \tag{1.4}$$

where $\widetilde{\mathbf{H}}^{1}(\boldsymbol{\Omega})$ is a subset of $\mathbf{H}^{1}$ of functions with zero mean.
3. Korn's inequality.

$$\int\limits_{\boldsymbol{\Omega}} \left(|\mathbf{v}|^2 + |\varepsilon(\mathbf{v})|^2\right) \mathbf{dx} \ge \mu_{\boldsymbol{\Omega}} \|\mathbf{v}\|^2_{1,2,\boldsymbol{\Omega}}, \ \forall \mathbf{v} \in \mathbf{H}^{1}(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}}), \tag{1.5}$$

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** $\mathbf{g}$ 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 $\mathbf{g}$ taken over the i-th variable.

Derivatives of $\mathbf{L^q}$–functions

Any function g from the space $\mathbf{L^q(\Omega)}$ $(\mathbf{q} \geq \mathbf{1})$ 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 $\mathbf{g}_{\alpha} \in \mathbf{L}^{\mathbf{q}}(\boldsymbol{\Omega})$.

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

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

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

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

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 $\mathbf{H^{-1}}(\mathbf{\Omega})$.

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

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

$$\boxed{\langle \mathbf{g}, \varphi \rangle \leq [\![\mathbf{g}]\!] \; \|\nabla \varphi\|_{\Omega}}$$

Generalized solutions

The concept of generalized solutions to PDE's came from **Petrov-Bubnov-Galerkin method**.

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

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.

References

1. I. G. Bubnov. *Selected Works.* Sudpromgiz, Leningrad (1956).
2. 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).
3. O. A. Ladyzhenskaya, *The boundary value problems of mathematical physics*. Springer-Verlag, New York, 1985 (in Russian 1970).
4. 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).

Definition

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

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

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

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

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

It has an inverse $\mathbf{A}^{-1} \in \mathcal{L}(\mathcal{V}, \mathcal{V})$, such that

$$\|\mathbf{A}\| \leq \mathbf{c_2}, \quad \|\mathbf{A}^{-1}\| \leq \frac{\mathbf{1}}{\mathbf{c_1}}$$

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

$$\frac{1}{2}\mathbf{B}(\mathbf{u}-\mathbf{v},\mathbf{u}-\mathbf{v}) = \frac{1}{2}\mathbf{B}(\mathbf{v},\mathbf{v}) - < \mathbf{f},\mathbf{v} > + \tag{1.6}$$

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

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

From this identity we derive two important results:

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

Let us show (a), i.e., that from (1.6) 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.

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:

$$\frac{1}{2}\mathbf{B}(\mathbf{u}-\mathbf{v},\mathbf{u}-\mathbf{v}) = \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}). \tag{1.7}$$

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

$$\boxed{\mathbf{B}(\mathbf{u}-\mathbf{u_h},\mathbf{u}-\mathbf{u_h}) = \inf_{\mathbf{v_h}\in\mathbf{K_h}} \mathbf{B}(\mathbf{u}-\mathbf{v_h},\mathbf{u}-\mathbf{v_h})} \tag{1.8}$$

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

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

for

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

we find that

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

provided that

- **Exact solution is $\mathbf{H^2}$ – regular;**
- **$\mathbf{u_h}$ 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 can hardly serve these purposes because, in general the constant C in such an estimate is either unknown or highly overestimated.
Therefore, a priori convergence estimates have mainly a theoretical value: they show that an approximation method is correct "in principle".

For these reasons, starting from late 70th 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 .

Lecture 2.
A CONCISE OVERVIEW OF A POSTERIORI ERROR ESTIMATION METHODS FOR APPROXIMATIONS OF DIFFERENTIAL EQUATIONS.

Lecture plan

- Heuristic Runge's rule;
- Prager and Synge estimate. Estimate of Mikhlin;
- Estimates using negative norm of the equation residual;
  - Basic idea;
  - Estimates in 1D case;
  - Estimates in 2D case;
  - Comments;
- Methods based on post–processing;
- Methods using adjoint problems;

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.

Carle Runge

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:

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

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

Also, the quantity $[u_h - u_{h_{ref}}]$ 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 $[\mathbf{u_h} - \mathbf{u_{h_{ref}}}]$ 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:

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

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)

W. Prager and J. L. Synge

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

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

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:

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

where

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

Dual problem

Since

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

where

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

we find that

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

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

$$\boxed{\frac{1}{2}\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \le \frac{1}{2}\|\nabla \mathbf{v} - \boldsymbol{\tau}\|^2, \qquad \forall \boldsymbol{\tau} \in \mathbf{Q_f}.} \tag{2.1}$$

Difficulties

Estimates of Prager and Synge and of Mikhlin are valid for any $\mathbf{v} \in \overset{\circ}{\mathbf{H}}_1(\mathbf{\Omega})$, 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.

Errors and Residuals: first glance

If an analyst is not sure in an approximate solution, then the very first idea that comes to his mind is to substitute it into the equation considered, i.e. to look at the equation residual.

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

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

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

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

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

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

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

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

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

$$\boxed{(\textbf{Cond}\,\mathcal{A})^{-1}\frac{\|\mathbf{r}\|}{\|\mathbf{f}\|} \leq \frac{\|\mathbf{e}\|}{\|\mathbf{u}\|} \leq \textbf{Cond}\,\mathcal{A}\,\frac{\|\mathbf{r}\|}{\|\mathbf{f}\|}.} \qquad (2.2)$$

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 (2.2).

Elliptic equations

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

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

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

$$\|\mathbf{v} - \mathbf{u}\|_{\mathbf{X}} \leq \mathbf{C}\|\mathcal{A}\mathbf{v} + \mathbf{f}\|_{\mathbf{Y}}, \tag{2.3}$$

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

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

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

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

which implies the **energy estimate**

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

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

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

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

$$\|\nabla(\mathbf{u} - \mathbf{v})\|_{2,\Omega} \leq \mathbf{C}_{\Omega}\|\mathbf{f} + \boldsymbol{\Delta}\mathbf{v}\|_{2,\Omega}, \tag{2.4}$$

whose right-hand side of (2.4) 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.,

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

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

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

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 $\langle \mathbf{f}, \mathbf{w} \rangle \leq [\![\, \mathbf{f} \,]\!] \, \|\nabla \mathbf{w}\|_{2,\Omega}$, we obtain another "energy estimate"

$$\|\nabla \mathbf{u}\|_{2,\Omega} \leq [\![\, \mathbf{f} \,]\!] .$$

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

$$\|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega} \leq [\![\, \mathbf{f}+\boldsymbol{\Delta}\mathbf{v} \,]\!]. \tag{2.5}$$

where

$$[\![\, \mathbf{f}+\boldsymbol{\Delta}\mathbf{v} \,]\!] = \sup_{\varphi\in\overset{\circ}{\mathbf{H}}{}^1(\Omega)} \frac{|\langle \mathbf{f}+\boldsymbol{\Delta}\mathbf{v},\varphi\rangle|}{\|\nabla\varphi\|} =$$

$$= \sup_{\varphi\in\overset{\circ}{\mathbf{H}}{}^1(\Omega)} \frac{|\int_\Omega \nabla(\mathbf{u}-\mathbf{v})\cdot\nabla\varphi|}{\|\nabla\varphi\|} \leq \sup_{\varphi\in\overset{\circ}{\mathbf{H}}{}^1(\Omega)} \frac{\|\nabla(\mathbf{u}-\mathbf{v})\|\|\nabla\varphi\|}{\|\nabla\varphi\|} \leq \|\nabla(\mathbf{u}-\mathbf{v})\|$$

Thus, for the problem considered

$$\|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega} = [\![\, \mathbf{f} + \boldsymbol{\Delta}\mathbf{v} \,]\!] \ !!! \tag{2.6}$$

From (2.6), it readily follows that

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

We observe that the estimate (2.6) is **consistent**.

Diffusion equation

Similar estimates can be derived for

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

where

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

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

$$\lambda_{\mathsf{min}}|\boldsymbol{\eta}|^{\mathbf{2}} \leq \mathbf{a_{ij}}(\mathbf{x})\boldsymbol{\eta}_{\mathbf{i}}\boldsymbol{\eta}_{\mathbf{j}} \leq \lambda_{\mathsf{max}}|\boldsymbol{\eta}|^{\mathbf{2}}, \quad \forall\boldsymbol{\eta} \in \mathbb{R}^{\mathbf{n}},\ \mathbf{x} \in \boldsymbol{\Omega},$$

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

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

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

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

$$\|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\Omega} \leq \lambda_{\min}^{-1} [\![\, \mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}) \,]\!] . \tag{2.7}$$

Next,

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

$$= \sup_{\varphi\in\overset{\circ}{\mathbf{H}}{}^1(\boldsymbol{\Omega})} \frac{|\,\int_{\boldsymbol{\Omega}} \mathbf{A}\nabla(\mathbf{u}-\mathbf{v})\cdot\nabla\varphi\,\mathbf{dx}\,|}{\|\nabla\varphi\|_{2,\boldsymbol{\Omega}}} \leq \lambda_{\max}\|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\boldsymbol{\Omega}}. \quad (2.8)$$

Combining (2.7) and (2.8) we obtain

$$\boxed{\lambda_{\max}^{-1} \, [\![\, \mathbf{R}(\mathbf{v}) \,]\!] \leq \|\nabla(\mathbf{u}-\mathbf{v})\|_{2,\boldsymbol{\Omega}} \leq \lambda_{\min}^{-1} \, [\![\, \mathbf{R}(\mathbf{v}) \,]\!] \,,} \quad (2.9)$$

where $\mathbf{R}(\mathbf{v}) = \mathbf{f} + \mathbf{div}\,(\mathbf{A}\nabla\mathbf{v}) \in \mathbf{H}^{-1}(\boldsymbol{\Omega})$. 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:

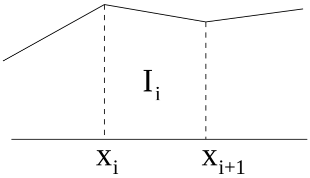
- Galerkin orthogonality property;
- $\mathbf{H^1} \longrightarrow \mathbf{V_h}$ interpolation estimates by Clément.

Explicit residual method in 1D case

Take the simplest model

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

Let $\mathbf{I} := (\mathbf{0}, \mathbf{1})$, $\mathbf{f} \in \mathbf{L}^2(\mathbf{I})$, $\alpha(\mathbf{x}) \in \mathbf{C}(\bar{\mathbf{I}}) \geq \alpha_0 > \mathbf{0}$. Divide $\mathbf{I}$ into a number of subintervals $\mathbf{I_i} = (\mathbf{x_i}, \mathbf{x_{i+1}})$, where $\mathbf{x_0} = \mathbf{0}$, $\mathbf{x_{N+1}} = \mathbf{1}$, and $|\mathbf{x_{i+1}} - \mathbf{x_i}| = \mathbf{h_i}$. Assume that $\mathbf{v} \in \overset{\circ}{\mathbf{H}}{}^1(\mathbf{I})$ and it is smooth on any interval I_i.



In this case,

$$[\![\mathbf{R}(\mathbf{v})]\!] = \sup_{\mathbf{w}\in \mathbf{V}_0(\mathbf{I}),\ \mathbf{w}\neq 0} \frac{\int_0^1 (-\alpha \mathbf{v}'\mathbf{w}' + \mathbf{f}\mathbf{w})\mathbf{dx}}{\|\mathbf{w}'\|_{2,\mathbf{I}}} =$$

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

$$= \sup_{\mathbf{w}\in \mathbf{V}_0(\mathbf{I}),\ \mathbf{w}\neq 0} \frac{\sum_{i=0}^{N} \int_{\mathbf{I}_i} \mathbf{r}_i(\mathbf{v})\mathbf{w}\,\mathbf{dx} + \sum_{i=1}^{N} \alpha(\mathbf{x}_i)\mathbf{w}(\mathbf{x}_i)\mathbf{j}(\mathbf{v}'(\mathbf{x}_i))}{\|\mathbf{w}'\|_{2,\mathbf{I}}},$$

where $\mathbf{j}(\phi(\mathbf{x})) := \phi(\mathbf{x}+\mathbf{0}) - \phi(\mathbf{x}-\mathbf{0})$ is the "jump–function" and $\mathbf{r}_i(\mathbf{v}) = (\alpha\mathbf{v}')' + \mathbf{f}$ 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

$$[\![\, \mathbf{R}(\mathbf{u_h}) \,]\!] = \sup_{\mathbf{w} \in \mathbf{V}_0(\mathbf{I})} \frac{\int_0^1 (-\alpha \mathbf{u}_h' (\mathbf{w} - \pi_h \mathbf{w})' + \mathbf{f}(\mathbf{w} - \pi_h \mathbf{w})) \, \mathbf{dx}}{\|\mathbf{w}'\|_{2,\mathbf{I}}},$$

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

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

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

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

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

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

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

$$\boxed{[\![\, R(u_h) \,]\!] \le \left(\sum_{i=0}^{N} c_i^2 \|r_i(u_h)\|_{2,I_i}^2 \right)^{1/2}.} \qquad (2.10)$$

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 $\mathbf{c_i}$ are easy to find. Indeed, let γ_i be a constant that satisfies the condition

$$\inf_{\mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^1(\mathbf{I_i})} \frac{\|\mathbf{w}'\|^2_{2,\mathbf{I_i}}}{\|\mathbf{w} - \pi_\mathbf{h} \mathbf{w}\|^2_{2,\mathbf{I_i}}} \geq \gamma_\mathbf{i}.$$

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

$$\|\mathbf{w} - \pi_\mathbf{h} \mathbf{w}\|_{2,\mathbf{I_i}} \leq \gamma_\mathbf{i}^{-1/2} \|\mathbf{w}'\|_{2,\mathbf{I_i}}$$

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

Let us estimate γ_{I_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 $(0, h)$.

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

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

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

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

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 $\mathbf{p_s}(\mathbf{x}) \in \mathbf{P^1}(\omega_\mathbf{s})$ 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 $\boldsymbol{\pi}_\mathbf{h}$ is defined by setting

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

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

$$\|\mathbf{v}-\pi_{\mathbf{h}}\mathbf{v}\|_{\mathbf{2},\mathbf{T_i}} \leq \mathbf{c_i^T}\,\mathbf{diam}\,(\mathbf{T_i})\|\mathbf{v}\|_{\mathbf{1},\mathbf{2},\omega_{\mathbf{N}}(\mathbf{T_i})}, \tag{2.11}$$

$$\|\mathbf{v}-\pi_{\mathbf{h}}\mathbf{v}\|_{\mathbf{2},\mathbf{E_{ij}}} \leq \mathbf{c_{ij}^E}|\mathbf{E_{ij}}|^{\mathbf{1/2}}\|\mathbf{v}\|_{\mathbf{1},\mathbf{2},\omega_{\mathbf{E}}(\mathbf{T_i})}, \tag{2.12}$$

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

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

Quotient relations for the constants

Evaluation of $\mathbf{c_i^T}$ and $\mathbf{c_{ij}^E}$ 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 $\boldsymbol{\sigma}_h = A\nabla u_h$. Then,

$$[\![R(u_h)]\!] = \sup_{w \in V_0} \frac{\int_\Omega (f(w - \pi_h w) - \boldsymbol{\sigma}_h \cdot \nabla (w - \pi_h w))\, dx}{\|\nabla w\|_{2,\Omega}}.$$

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

$$\int_{T_i} \boldsymbol{\sigma}_h \cdot \nabla (w - \pi_h w)\, dx =$$

$$= \sum_{E_{ij} \subset \partial T_i} \int_{E_{ij}} (\boldsymbol{\sigma}_h \cdot \boldsymbol{\nu})(w - \pi_h w) ds - \int_{T_i} \operatorname{div} \boldsymbol{\sigma}_h (w - \pi_h w)\, dx,$$

Since on the boundary $w - \pi_h w = 0$, we obtain

$$[\![R(u_h)]\!] = \sup_{w \in V_0} \left\{ \frac{\sum_{i=1}^N \int_{T_i} (\operatorname{div} \boldsymbol{\sigma}_h + f)(w - \pi_h w)\, dx}{\|\nabla w\|_{2,\Omega}} + \right.$$

$$\left. + \frac{\sum_{i=1}^N \sum_{j>i}^N \int_{E_{ij}} j(\boldsymbol{\sigma}_h \cdot \boldsymbol{\nu}_{ij})(w - \pi_h w)\, ds}{\|\nabla w\|_{2,\Omega}} \right\}.$$

First term in sup

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

Then, the first sum is estimated as follows:

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

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

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

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

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

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

$$[\![\, R(u_h) \,]\!] \leq C_0 \Bigg(\Big(\sum_{i=1}^{N} (c_i^T)^2 \operatorname{diam}(T_i)^2 \| \operatorname{div} \sigma_h + f \|_{2,T_i}^2 \Big)^{1/2} + \\ + \Big(\sum_{i=1}^{N} \sum_{j>i}^{N} (c_{ij}^E)^2 \, |E_{ij}| \, \| j(\sigma_h \cdot \nu_{ij}) \|_{2,E_{ij}}^2 \Big)^{1/2} \Bigg). \quad (2.13)$$

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 $\mathbf{c}_i^T$ have approximately the same value and can be replaced by a single constant $\mathbf{c}_1$. If the constants $\mathbf{c}_{ij}^E$ are also estimated by a single constant $\mathbf{c}_2$, then we have

$$[\![\mathbf{R}(\mathbf{u}_h)]\!] \leq \mathbf{C} \left(\sum_{i=1}^{N} \eta^2(\mathbf{T}_i) \right)^{1/2}, \tag{2.14}$$

where $\mathbf{C} = \mathbf{C}(\mathbf{c}_1, \mathbf{c}_2, \mathbf{C}_0)$ and

$$\eta^2(\mathbf{T}_i) = \mathbf{c}_1^2 \mathbf{diam}\,(\mathbf{T}_i)^2 \|\mathbf{div}\,\sigma_h + \mathbf{f}\|_{2,\mathbf{T}_i}^2 + \frac{\mathbf{c}_2^2}{2} \sum_{\mathbf{E}_{ij} \subset \partial \mathbf{T}_i} |\mathbf{E}_{ij}| \|\mathbf{j}(\sigma_h \cdot \nu_{ij})\|_{2,\mathbf{E}_{ij}}^2 .$$

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

Comment 1

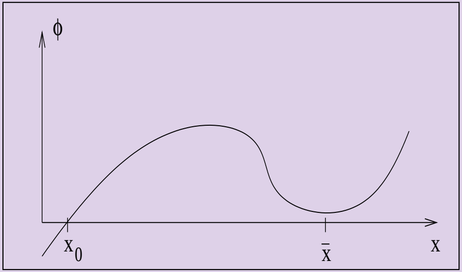
General form of the residual type a posteriori error estimates is as follows:

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

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. Babuška and W. C. Rheinboldt. A-posteriori error estimates for the finite element method. Internat. J. Numer. Meth. Engrg., 12(1978).
I. Babuška 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.
Finding the collection of sharp constants $\mathbf{c_i}$ presents a special and often not an easy problem: see, e.g.,
C. Carstensen and S. A. Funken. Costants in Clement's–interpolation error and residual based a posteriori error estimates in finite element methods. *East–West J. Numer. Anal.* 8(2000), N3.

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 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

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

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

Ideal estimator has $\mathbf{i}_{\mathbf{eff}} = \mathbf{1}$. However, in real life situations it is hardly possible, so that values $\mathbf{i}_{\mathbf{eff}}$ 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

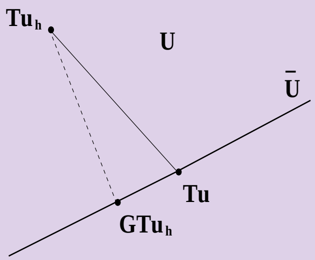
The estimator $\mathcal{E}$ is called **asymptotically equivalent to the error** if for a sequence of approximate solutions $\{u_h\}$ obtained on consequently refined meshes there holds the relation

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

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 $\mathbb{G}(\mathrm{T}\mathbf{v_h}) \in \overline{\mathbf{U}}$, $\forall \mathbf{v_h} \in \mathbf{V_h}$. 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 $\alpha < 1$ such that for the mapping T the estimate

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

Two–sided estimate

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

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

Thus, if $\boldsymbol{\alpha} << 1$, 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.,

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

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:

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

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

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

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

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

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

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

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

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

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

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

Various averaging formulas of this type are represented in the form

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

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

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

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. 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.

Discrete averaging on patches

Consider the problem

$$\inf_{\mathbf{g}\in\mathbb{P}^{\mathbf{k}}(\boldsymbol{\Omega}_{\mathbf{i}})} \sum_{\mathbf{s}=1}^{\mathbf{m_i}} |\mathbf{g}(\mathbf{x_s}) - \mathrm{T}\mathbf{u_h}(\mathbf{x_s})|^2 \,,$$

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

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

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 $\bar{\mathbf{g}}_\mathbf{h} \in \mathbf{V_h}(\boldsymbol{\Omega}) \subset \overline{\mathbf{U}}$ such that

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

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.,

C. Carstensen, S. Bartels. 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)

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

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

However, it was discovered see, e.g.,
L. A. Oganesjan and L. A. Ruchovec. *Z. Vychisl. 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

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

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

M. Křižek 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 α 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. 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).

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

$$\boldsymbol{\Lambda}^{\star}\mathrm{T}u + f = 0, \qquad \mathrm{Tu} = \mathcal{A}\boldsymbol{\Lambda}u, \tag{2.15}$$

where $\mathcal{A}$ is a positive definite operator, $\boldsymbol{\Lambda}$ is a linear continuous operator, and $\boldsymbol{\Lambda}^{\star}$ is the adjoint operator.
In continuum mechanics, equations of the type (2.15) 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.,

$$< \ell, \mathbf{v} - \mathbf{u} >= \int_{\boldsymbol{\Omega}} \varphi_{\mathbf{0}} \left(\mathbf{v} - \mathbf{u}\right) \mathbf{dx},$$

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

$$\mathbf{Au} = \mathbf{f},$$

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,

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

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

$$< \ell, \mathbf{u} - \mathbf{v} > = < \mathbf{f} - \mathbf{A}\mathbf{v}, \mathbf{u}_\ell > . \tag{2.16}$$

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

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

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

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

Since

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

and

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

We arrive at the relation

$$\int_\Omega \ell(u_f - u_{fh})dx = \int_\Omega A\nabla(u_\ell - u_{\ell h}) \cdot \nabla(u_f - u_{fh})dx \quad (2.17)$$

whose right-hand side is expressed in the form

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

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

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

This relation implies the estimate

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

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

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

$$+\frac{1}{2} \, \|\mathbf{j}(\nu_i \cdot \mathbf{A}\nabla\mathbf{u_{fh}})\|_{2,\partial T_i} \, \|\mathbf{u}_\ell - \mathbf{u}_{\ell h}\|_{2,\partial T_i} \Big\}.$$

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

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 $h^{\boldsymbol{\alpha}}|u_\ell|_{2,2,T_i}$ with $\boldsymbol{\alpha} = 1$ 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,

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

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

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

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

Integrating by parts, we obtain

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

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., $\|\mathbf{u}_\ell\| \leq \mathbf{C_1} \|\ell\|$), then we obtain

$$\|\mathbf{u}_\ell - \pi_\mathbf{h}(\mathbf{u}_\ell)\|_{\mathbf{T_i}} \leq \mathbf{C_1} \mathbf{h}^{\boldsymbol{\alpha}_1} \|\ell\|, \qquad \|\mathbf{u}_\ell - \pi_\mathbf{h}(\mathbf{u}_\ell)\|_{\partial \mathbf{T_i}} \leq \mathbf{C_1} \mathbf{h}^{\boldsymbol{\alpha}_2} \|\ell\|$$

with certain $\boldsymbol{\alpha}_\mathbf{k}$.
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 $\mathbf{C_1} \mathbf{h}^{\boldsymbol{\alpha}_1}$ and $\mathbf{C_2} \mathbf{h}^{\boldsymbol{\alpha}_2}$.

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

We end this lecture with one comment concerning the terminology In the existing 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 that is behind this terminology may be quite different because the word *"duality"* is used in at least 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.
FUNCTIONAL A POSTERIORI ESTIMATES. FIRST EXAMPLES.

Lecture goal

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

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

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.

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:

$$\mathbf{\Phi}(\mathbf{u}-\mathbf{v}) \leq \mathbf{M}\left(\mathcal{D}, \mathbf{v}\right) \qquad \forall \mathbf{v} \in \mathbf{V}\,! \tag{3.1}$$

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

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'

These results are summarized in S. Repin. *Math. Comput.*, 2000.

Nonvariational method 2000'

see S. Repin. *Proc. St.-Petersburg Math. Society*, 2001.
Complete list of publications on the matter can be found in the References appended to the Lectures.

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

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.

Variational Method

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

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

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

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

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

where

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

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

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

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

$$\inf_{\mathbf{v}\in\mathbf{V_0}}\int_\Omega\left(\nabla\mathbf{v}\cdot\mathbf{y}-\frac{1}{2}|\mathbf{y}|^2-\mathbf{f}\mathbf{v}\right)\mathbf{dx} = -\frac{1}{2}\|\mathbf{y}\|^2+\inf_{\mathbf{v}\in\mathbf{V_0}}\int_\Omega(\nabla\mathbf{v}\cdot\mathbf{y}-\mathbf{f}\mathbf{v})\mathbf{dx} =$$
$$= \begin{cases} -\frac{1}{2}\|\mathbf{y}\|^2 & \text{if } \mathbf{y}\in\mathbf{Q_f} := \{\mathbf{y}\in\mathbf{Y}\,|\,\mathbf{div y}+\mathbf{f}=\mathbf{0}\} \\ -\infty & \text{if } \mathbf{y}\notin\mathbf{Q_f} \end{cases}$$

Dual Problem

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

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

where

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

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*

$$\sup_{\mathbf{y}\in\mathbf{Y}} \inf_{\mathbf{x}\in\mathbf{X}} \mathbf{L}(\mathbf{x},\mathbf{y}) \leq \inf_{\mathbf{x}\in\mathbf{X}} \sup_{\mathbf{y}\in\mathbf{Y}} \mathbf{L}(\mathbf{x},\mathbf{y}). \tag{3.2}$$

Proof

It is easy to see that

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

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

proof

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

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

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

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

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

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

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

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

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

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

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

Now, we have

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

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

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

How to estimate the above infimum?

Various methods give one and the same answer:

$$\inf_{\mathbf{q}\in\mathbf{Q_f}} \|\mathbf{y}-\mathbf{q}\| \leq [\![\,\mathbf{div y}+\mathbf{f}\,]\!] \qquad \mathbf{y} \in \mathbf{L}^2(\boldsymbol{\Omega}), \tag{3.3}$$

$$\inf_{\mathbf{q}\in\mathbf{Q_f}} \|\mathbf{y}-\mathbf{q}\| \leq \mathbf{C}_{\boldsymbol{\Omega}}\|\mathbf{div y}+\mathbf{f}\| \qquad \mathbf{y} \in \mathbf{H}(\boldsymbol{\Omega},\mathbf{div}), \tag{3.4}$$

Proof

To prove these estimates we consider an auxiliary problem

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

$$\int_{\boldsymbol{\Omega}} \nabla\boldsymbol{\eta} \cdot \nabla \mathbf{w} \mathbf{dx} = \int_{\boldsymbol{\Omega}} (\mathbf{f} + \mathrm{div}\mathbf{y})\mathbf{w}\mathbf{dx} = \int_{\boldsymbol{\Omega}} (\mathbf{fw} - \mathbf{y} \cdot \nabla\mathbf{w})\mathbf{dx}$$

$$\int_{\boldsymbol{\Omega}} \overbrace{(\nabla\boldsymbol{\eta} + \mathbf{y})}^{\bar{\mathbf{q}}} \cdot \nabla\mathbf{w}\,\mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{fw}\,\mathbf{dx} \quad \forall\mathbf{w} \in \mathbf{V_0}$$

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

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

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

Then

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

Here

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

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

If **y** has a square summable divergence, then we have

$$[\![\mathbf{div} \mathbf{y} + \mathbf{f}]\!] = \sup_{\mathbf{w} \in \mathbf{V}_0} \frac{\int_\Omega (\mathbf{div} \mathbf{y} + \mathbf{f}) \mathbf{w} d\mathbf{x}}{\|\nabla \mathbf{w}\|} \leq \mathbf{C}_\Omega \|\mathbf{div} \mathbf{y} + \mathbf{f}\|,$$

where $\mathbf{C}_\Omega$ is the constant in the Friederichs–Steklov inequality for the domain $\mathbf{\Omega}$. Thus, **by taking the flux vector–valued function in the space that contains the flux of the true solution we make a "noncomputable" negative norm computable**.

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

$$\|\nabla(\mathbf{v}-\mathbf{u})\| \leq \|\nabla \mathbf{v}-\mathbf{y}\| + \inf_{\mathbf{q}\in \mathbf{Q_f}} \|\mathbf{y}-\mathbf{q}\| \leq$$
$$\|\nabla \mathbf{v}-\mathbf{y}\| + \mathbf{C}_{\boldsymbol{\Omega}}\|\mathbf{div}\mathbf{y}+\mathbf{f}\|.$$

Above presented *modus operandi* can be viewed as a simplest version of the variational approach to the derivation of Functional Error Majorants.

Method of integral identities. First glance.

For many problems the variational techniques cannot be applied because they 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. Later this method was applied to parabolic problems:

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

Nonvariational method in the simplest case

Let us expose its simplest version adapted to our model problem.
We have seen that

$$\|\nabla(\mathbf{u}-\mathbf{v})\| \leq [\![\, \mathbf{\Delta v}+\mathbf{f}\,]\!]$$

Instead of the estimation of the negative norm by Galerkin orthogonality and special intepolation estimates we suggest **another method** of finding an upper bound that is based on the functional relation

$$\int_\Omega (\mathbf{divyw} + \nabla \mathbf{w}\cdot\mathbf{y})\,\mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{w} \in \mathbf{V_0}$$

We have

$$
[\![\, \Delta \mathbf{v} + \mathbf{f} \,]\!] = \sup_{\mathbf{w}\in \mathbf{V}_0} \frac{\int_\Omega (\nabla \mathbf{v} \cdot \nabla \mathbf{w} - \mathbf{f}\mathbf{w})}{\|\nabla \mathbf{w}\|} =
$$

$$
\sup_{\mathbf{w}\in \mathbf{V}_0} \frac{\int_\Omega (\nabla \mathbf{v} \cdot \nabla \mathbf{w} - \mathbf{f}\mathbf{w} - (\mathbf{div}\mathbf{y}\,\mathbf{w} + \nabla \mathbf{w} \cdot \mathbf{y}))}{\|\nabla \mathbf{w}\|} =
$$

$$
\sup_{\mathbf{w}\in \mathbf{V}_0} \frac{\int_\Omega ((\nabla \mathbf{v} - \mathbf{y}) \cdot \nabla \mathbf{w} - (\mathbf{f} + \mathbf{div}\mathbf{y})\mathbf{w})\mathbf{dx}}{\|\nabla \mathbf{w}\|} =
$$

$$
\sup_{\mathbf{w}\in \mathbf{V}_0} \frac{\|\nabla \mathbf{v} - \mathbf{y}\|\|\nabla \mathbf{w}\| + \|\mathbf{f} + \mathbf{div}\mathbf{y}\|\|\mathbf{w}\|}{\|\nabla \mathbf{w}\|} \leq
$$

$$
\leq \|\nabla \mathbf{v} - \mathbf{y}\| + \mathbf{C}_\Omega \|\mathbf{f} + \mathbf{div}\mathbf{y}\|.
$$

Functional error estimate. Meaning and properties

For the problem

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

we have obtained the estimate

$$\boxed{\|\nabla(\mathbf{u} - \mathbf{v})\| \leq \|\nabla\mathbf{v} - \mathbf{y}\| + \mathbf{C}_{\boldsymbol{\Omega}}\|\mathbf{divy} + \mathbf{f}\|} \qquad (3.5)$$

The estimate is valid for any $\mathbf{v} \in \mathbf{V}_0$ and $\mathbf{y} \in \mathbf{H}(\boldsymbol{\Omega}, \mathbf{div})$
Two terms in the right–hand side have a clear sense: they **present measures of the errors in two basic relations**

$$\mathbf{p} = \nabla\mathbf{u}, \qquad \mathbf{divp} + \mathbf{f} = \mathbf{0} \quad \text{in } \boldsymbol{\Omega}$$

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

$$\|\nabla \mathbf{u}\| \;\leq\; \mathbf{C}_{\boldsymbol{\Omega}} \|\mathbf{f}\|.$$

Therefore, no constant less than $\mathbf{C}_{\boldsymbol{\Omega}}$ 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 (3.5) 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

$$\mathbf{2ab} \leq \beta \mathbf{a}^2 + \frac{\mathbf{1}}{\beta}\mathbf{b}^2, \qquad \beta > \mathbf{0}$$

we rewrite this estimate in the form

$$\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \leq \tag{3.6}$$

$$\leq (\mathbf{1}+\beta)\|\nabla\mathbf{v}-\mathbf{y}\|^2 + \frac{\mathbf{1}+\beta}{\beta}\mathbf{C}_\Omega^2\|\mathbf{div}\mathbf{y}+\mathbf{f}\|^2$$

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 $\mathbf{y}$.

Deviation Majorant

Denote the right–hand side of (3.6) by $\mathcal{M}_{\oplus}$, i.e.,

$$\mathcal{M}_{\oplus}(\mathbf{v},\mathbf{y},\beta,\mathbf{C}_{\Omega},\mathbf{f}) := (\mathbf{1}+\beta)\|\nabla\mathbf{v}-\mathbf{y}\|^{2} + \frac{\mathbf{1}+\beta}{\beta}\mathbf{C}_{\Omega}^{2}\|\mathbf{div}\mathbf{y}+\mathbf{f}\|^{2}.$$

This functional provides an upper bound for the norm of the deviation of **v** from **u**. Therefore, it is natural to call it the **Deviation Majorant**.

BVP $\Delta u + f = 0$ has another variational formulation

$$\boxed{\inf_{\substack{\mathbf{v}\in \mathbf{V_0},\\ \beta>0,\\ \mathbf{y}\in \mathbf{H}(\boldsymbol{\Omega},\mathrm{div}),}} \mathcal{M}_{\oplus}(\mathbf{v}, \mathbf{y}, \beta, \mathbf{C}_{\boldsymbol{\Omega}}, \mathbf{f})}$$

- 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 $\mathbf{C}_{\boldsymbol{\Omega}}$, which is problem independent;

In principle, one can select certain sequences of subspaces $\{\mathbf{V_{hk}}\} \in \mathbf{V_0}$ and $\{\mathbf{Y_{hk}}\} \in \mathbf{H}(\mathbf{\Omega}, \mathbf{div})$ and minimize the Error Majorant with respect to these subspaces

$$\inf_{\substack{\mathbf{v}\in\mathbf{V_{hk}},\\ \boldsymbol{\beta}>0,\\ \mathbf{y}\in\mathbf{Y_{hk}},}} \mathcal{M}_{\oplus}(\mathbf{v}, \mathbf{y}, \beta, \mathbf{C_\Omega}, \mathbf{f})$$

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

$$\gamma_{\mathbf{k}} := \inf_{\boldsymbol{\beta}>\mathbf{0}} \mathcal{M}_{\oplus}(\mathbf{v_k}, \mathbf{y_k}, \beta, \mathbf{C_\Omega}, \mathbf{f}) \longrightarrow \mathbf{0}$$

How to use the Majorants in practice?

Consider **CONFORMING FEM APPROXIMATIONS**.

We have 3 basic ways to use the deviation estimate:

(a) Use flux averaging on the mesh $\mathcal{T}_h$);

(b) Use flux averaging on the refined mesh $\mathbf{h_{ref}}$);

(c) Minimization with respect to $\mathbf{y}$.

(a) Use recovered gradients

Let $\mathbf{u_h} \in \mathbf{V_h}$, then

$$\mathbf{p_h} := \nabla \mathbf{u_h} \in \mathbf{L_2}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}}), \quad \mathbf{p_h} \notin \mathbf{H}(\mathbf{\Omega}, \mathrm{div}).$$

Use an averaging operator $\mathbf{G_h} : \mathbf{L_2}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}}) \to \mathbf{H}(\mathbf{\Omega}, \mathrm{div})$ and have a **directly computable estimate**

$$\|\nabla(\mathbf{u} - \mathbf{u_h})\| \leq \|\nabla \mathbf{u_h} - \mathbf{G_h p_h}\| + \mathbf{C_\Omega} \, \|\mathbf{div} \mathbf{G_h p_h} + \mathbf{f}\|$$

(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

$$\|\nabla(\mathbf{u}-\mathbf{u_h})\|^2 \leq \inf_{\mathbf{y_h}\in\mathbf{Y_h}}\Big\{(\mathbf{1}+\beta)\|\nabla\mathbf{u_h}-\mathbf{y_h}\| + \\ + \mathbf{C_\Omega}\left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\|\mathbf{div}\mathbf{y_h}+\mathbf{f}\|^2\Big\}$$

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

$$(\alpha(x)\, u')' = f(x),$$
$$u(a) = 0, \quad u(b) = u_b.$$

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

$$\mathcal{M}_{\oplus}(\mathbf{v},\beta,\mathbf{y})=(1+\beta)\left(\int\limits_a^b |\,\alpha\mathbf{v}'-\mathbf{y}\,|^2\,\mathbf{dx}+\frac{\mathbf{C}^2_{(a,b)}}{\beta}\int\limits_a^b|\mathbf{y}'-\mathbf{f}|^2\right)\mathbf{dx}. \quad (3.7)$$

In this simple model, **u** can be presented in the form

$$\mathbf{u}(\mathbf{x}) \;=\; \int_a^x \frac{1}{\alpha(\mathbf{t})}\int_a^t \mathbf{f}(\mathbf{z})\mathbf{dzdt} + \frac{\mathbf{x}}{\mathbf{b}}\left(\mathbf{u_b} - \int_a^b \frac{1}{\alpha(\mathbf{t})}\int_a^t \mathbf{f}(\mathbf{z})\mathbf{dzdt}\right).$$

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

$$\mathbf{G}(\mathbf{v}')(\mathbf{x_i}) = \frac{1}{2}(\mathbf{v}'(\mathbf{x_i} - \mathbf{0}) + \mathbf{v}'(\mathbf{x_i} + \mathbf{0})),$$

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 1

Let $\alpha(x) = 1$, $\mathbf{f(x) = c}$, $\mathbf{a = 0}$, $\mathbf{b = 1}$, $\mathbf{u_b = 1}$, e.g., we consider the problem

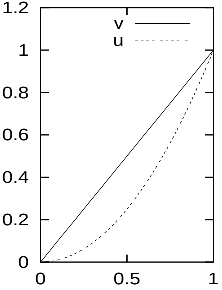
$$\mathbf{u'' = 2}, \quad \mathbf{u(0) = 0},\ \mathbf{u(1) = 1}.$$

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

$$\|\mathbf{(u - v)'}\|^2 = \int_0^1 \mathbf{c^2(x - 0.5)^2 dx = c^2/12 \approx 0.083c^2}.$$



Exact solution and an approximation.

Various y give different upper bounds

(a) Take $\mathbf{y} = \mathbf{v}' = \mathbf{1}$, then the first term in

$$\mathcal{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{y}) = (\mathbf{1} + \beta)\left(\int_0^1 |\, \mathbf{v}' - \mathbf{y}\,|^2\, \mathbf{dx} + \frac{\mathbf{1}}{\pi^2 \beta}\int_0^1 |\mathbf{y}' - \mathbf{f}|^2\right) \mathbf{dx}.$$

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

$$\mathcal{M}_{\oplus} = \frac{\mathbf{1}}{\pi^{\mathbf{2}}} \int_{\mathbf{0}}^{\mathbf{1}} \mathbf{c}^{\mathbf{2}} \mathbf{dx}.$$

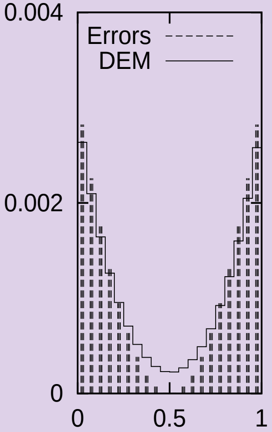
However, the integrand of the Majorant is a constant function, but the error is distributed in accordance with a parabolic law:

$$(\mathbf{u} - \mathbf{v})' = \mathbf{c}(\mathbf{x} - \mathbf{0.5})^{\mathbf{2}}.$$

(b). Take $\mathbf{y} = \mathbf{cx} + \mathbf{1} - \mathbf{c}/\mathbf{2}$. Then, $\mathbf{y}' = \mathbf{c}$ and the second term of the majorant vanishes. We have (for $\beta = 0$)

$$\mathcal{M}_{\oplus} = \int_{0}^{1} \mathbf{c}^2(\mathbf{x} - \mathbf{1}/\mathbf{2})^2 \mathbf{dx} = \mathbf{c}^2/\mathbf{12}.$$

We observe that both the global error and the error distribution are exactly reproduced. In real life computations such an "ideal" function **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 **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 φ 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.

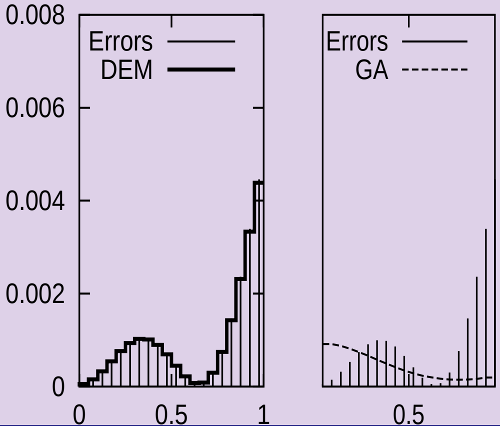
We take $\varphi = \mathbf{x}\sin(\pi\mathbf{x})$ and $\delta = 0.1,\ 0.01,\ 0.001$, and 0.

Table:

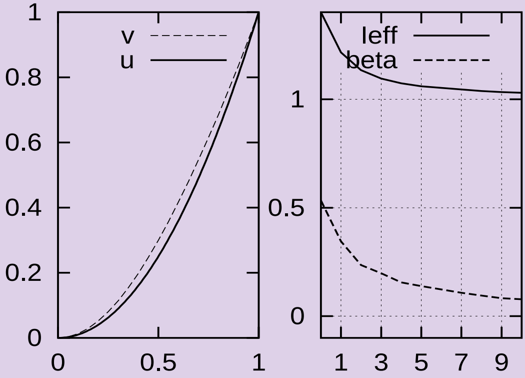
δ	e	$2\mathcal{M}_\oplus$	$2\mathcal{M}_\ominus$	$\mathrm{i}_{\mathrm{eff}}$	$\mathrm{i}_{\mathrm{esh}}$
0.1	0.019692	0.019743	0.019683	1.003	1.018
0.01	0.001022	0.001025	0.001013	1.003	1.011
0.001	0.000835	0.000839	0.000827	1.005	1.002
0	0.000833	0.000836	0.000825	1.004	1.002

In this experiment the Majorant was computed for $\frac{1}{2}\|\mathbf{e}\|^2$.

Error estimation for $\delta = 0.1$

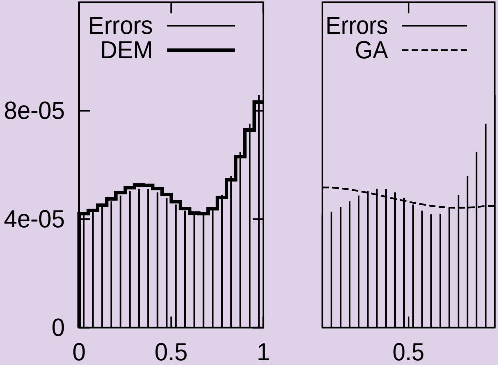


Functions u, v and i_{eff} for $\delta = 0.1$

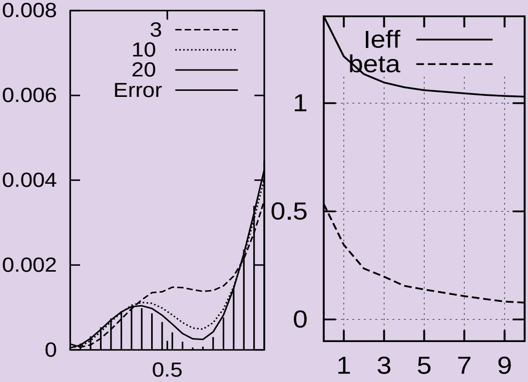


Error estimation for $\delta = 0.01$

A more precise approximation.

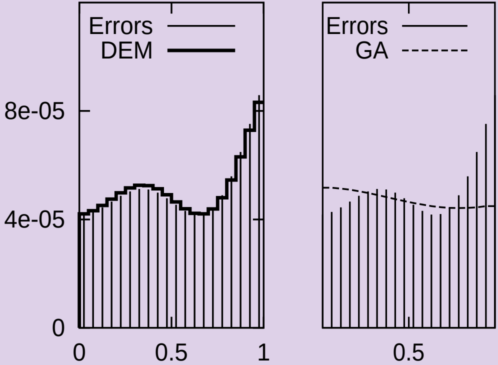


Functions e(y), β and $\mathbf{i}_{\text{eff}}$ for $\delta = 0.1$

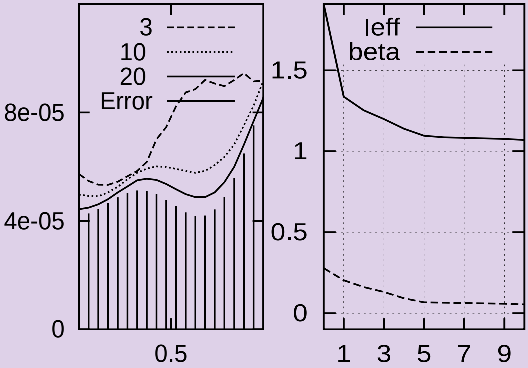


Error estimation for $\delta = 0.01$

A more precise approximation.

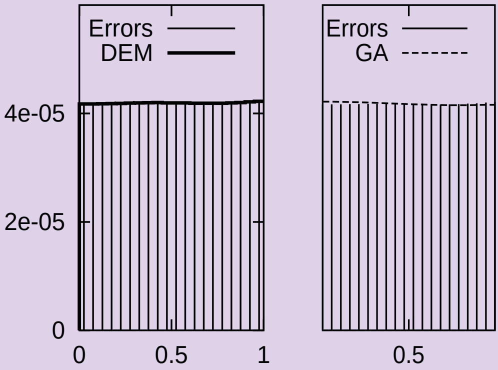


Functions e(y), β and $\mathbf{i}_{\text{eff}}$ for $\delta = 0.01$

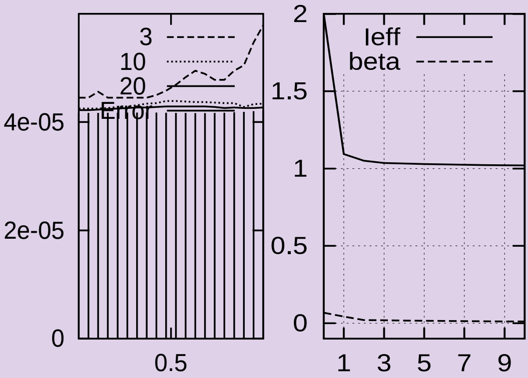


Error estimation for $\delta = 0.001$

Sharp approximation.

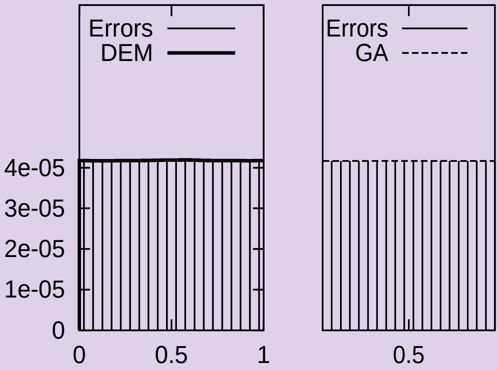


Functions e(y), β and $\mathbf{i}_{\mathrm{eff}}$ for $\delta = 0.001$

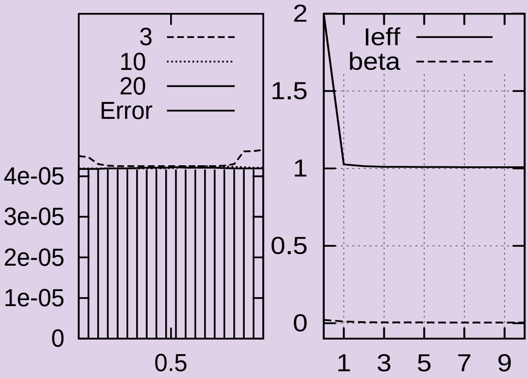


Error estimation for $\delta = 0$

Interpolant of the exact solution.



Functions e(y), β and $\mathbf{i}_{\mathrm{eff}}$ for $\delta = 0$



Functional a posteriori error estimates were derived by the methods of duality theory in convex analysis in 1996. These results are published in [7,8]. In [5,9,23,24], they were applied to certain linear and nonlinear variational problems with convex functionals. First consequent study of their computational properties was presented in [10]. Later a detailed investigation of the practical aspects was done in [1,3,18]. General a posteriori estimates for the class of convex functionals are derived and discussed in [3,5,11,12]. A posteriori estimates for a class of nonconvex problems are can be found in [6]. A posteriori estimates which take into account errors in main boundary conditions were derived in [20], there readers can also find a method of the derivation of the estimates based upon the orthogonal decomposition of the space L^2 (Helmgholts decomposition). In, [17,21,22] a posteriori estimates were derived for modeling errors in dimension reduction models. Estimates for the Stokes problem will be further discussed in this lecture course (see [15,16]). In [19], functional type a posteriori estimates were obtained for the Reissner-Mindlin plate.

1 . Frolov, P. Neittaanmäk and S. Repin. On computational properties of a posteriori error estimates based upon the method of duality error majorants. In *Proc. 5th European Conference on Numerical Mathematics and applications, Prague 2004*, 346–357.

2 . Gaevskaya and S. Repin. A posteriori error estimates for approximate solutions of linear parabolic problems. *Differential Equations*, 41 (2005), 7, 970–983.

3 . Neittaanmäki and S. Repin. *Reliable methods for computer simulation. Error control and a posteriori estimates*, Elsevier, New York, 2004.

4 . 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.

5 . 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).

6 . 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).

7 . Repin. A posteriori error estimation for nonlinear variational problems by duality theory. *Zapiski Nauchnych Seminarov POMI*, 243(1997) 201-214.

8 . Repin. A posteriori error estimation for variational problems with power growth functionals based on duality theory, *Zapiski Nauchnych Seminarov POMI*, 249(1997), 244-255.

9 . 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.

10 . Repin. A unified approach to a posteriori error estimation based on duality error majorants, *Mathematics and Computers in Simulation*, 50(1999), 313-329.

11 . Repin. A posteriori error estimation for variational problems with uniformly convex functionals, *Math. Comput.*, 69(230), 2000, 481-500.

12 . 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)*)

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

14 . 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.

15 . Repin. Aposteriori estimates for the Stokes problem, *Journal of Math. Sciences* 109 (2002), 5, 1950-1964.

16 . Repin. Estimates of deviations for generalized Newtonian fluids, *Zapiski Nauchn. Semin. V.A. Steklov Mathematical Institute in St.-Petersburg*

(POMI), 288(2002), 178-203.

17 . Repin. Estimates for errors in two-dimensional models of elasticity theory, *J. Math. Sci. (New York)*, 106 (2001), no. 3, 3027-3041.

18 . 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).

19 epin, S. I.; Frolov, M. E. An estimate for deviations from the exact solution of the Reissner-Mindlin plate problem. (Russian) Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 310 (2004), 145–157, 228.

20 . 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.

21 . Repin, S. Sauter and A. Smolianski. Duality Based A Posteriori Error estimator for the Dirichlet Problem, *Proc.Appl. Math. Mech.*, 2 (2003), 513-514.

22 . 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).

23 . 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.

24 . 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.

Lecture 4.
AN INTRODUCTION TO DUALITY THEORY.

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

$$\mathbf{div\, A \nabla u + f = 0}$$

with certain boundary conditions.
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.

Many problems in continuum mechanics encompassed in the general scheme are: **linear elasticity,**
biharmonic problems,
Kirhghoff and Mindlin plates,
deformation theory of elastoplasticity,
Stokes problem.
Also, this scheme is applicable to the **p-Laplace** equation and certain **nonlinear models in the theory of viscous fluids**.

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.

Lecture plan

- **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

$$\langle \boldsymbol{\xi}^*, \boldsymbol{\xi} \rangle = \boldsymbol{\xi}^* \cdot \boldsymbol{\xi} = \xi_i^* \xi_i .$$

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

$$\mathbf{G}(\mathbf{v}) := \int_{\Omega} \mathbf{g}(\mathbf{\Lambda v})\mathbf{dx}$$

where **g** is the "internal energy" or "dissipative potential", then the so–called "complementary energy" is given by the integral functional

$$\mathbf{G}^*(\mathbf{y}^*) := \int_{\Omega} \mathbf{g}^*(\mathbf{y}^*)\mathbf{dx},$$

where $\mathbf{g}^*$ is conjugate to **g** in the algebraic sense.

Example 1 (Diffusion problems)

Let $\mathbf{A} = \{\mathbf{a_{ij}}\}$ be a real, positive definite matrix and

$$\mathbf{g}(\boldsymbol{\xi}) = \frac{1}{2}\mathbf{A}\boldsymbol{\xi} \cdot \boldsymbol{\xi} = \frac{1}{2}\mathbf{a_{ij}}\boldsymbol{\xi}_\mathbf{i}\boldsymbol{\xi}_\mathbf{j}.$$

Then

$$\mathbf{g}^*(\boldsymbol{\xi}^*) = \sup_{\boldsymbol{\xi} \in \mathbf{E^d}} \left\{ \boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - \frac{1}{2}\mathbf{A}\boldsymbol{\xi} \cdot \boldsymbol{\xi} \right\}.$$

This supremum is attained on an element $\boldsymbol{\xi}_0$ such that

$$\boldsymbol{\xi}^* = \mathbf{A}\boldsymbol{\xi}_\mathbf{0} \implies \boldsymbol{\xi}_\mathbf{0} = \mathbf{A}^{-\mathbf{1}}\boldsymbol{\xi}^*.$$

Therefore, we have a pair of mutually conjugate functionals

$$\mathbf{g}(\boldsymbol{\xi}) = \frac{1}{2}\mathbf{A}\boldsymbol{\xi} \cdot \boldsymbol{\xi} \quad \text{and} \quad \mathbf{g}^*(\boldsymbol{\xi}^*) = \frac{1}{2}\mathbf{A}^{-\mathbf{1}}\boldsymbol{\xi}^* \cdot \boldsymbol{\xi}^*.$$

In diffusion type boundary–value problems we arrive at the functional (with $\mathbf{y} = \nabla \mathbf{v}$)

$$\frac{1}{2}\int_\Omega \mathbf{A}\mathbf{y}\cdot\mathbf{y}\,\mathbf{dx} \quad \mathbf{y}\in \mathbf{L}^2(\Omega,\mathbb{R}^{\mathbf{n}}),$$

which is mutually dual to

$$\frac{1}{2}\int_\Omega \mathbf{A}^{-1}\mathbf{y}^*\cdot\mathbf{y}^*\,\mathbf{dx} \quad \mathbf{y}^*\in \mathbf{L}^2(\Omega,\mathbb{R}^{\mathbf{n}})$$

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

$$\mathbf{g}^*(\varepsilon^*) = \sup_{\boldsymbol{\varepsilon} \in \mathbf{M}^{\mathbf{d}\times\mathbf{d}}} \left\{ \varepsilon^* : \varepsilon - \frac{1}{2}\mathbf{A}\varepsilon : \varepsilon \right\}.$$

This supremum is attained on an element ε_0 such that

$$\boldsymbol{\tau}^* = \mathbf{L}\varepsilon_{\mathbf{0}} \implies \varepsilon_{\mathbf{0}} = \mathbf{L}^{-\mathbf{1}}\varepsilon^*.$$

Therefore, we have a pair of mutually dual functionals

$$\mathbf{g}(\varepsilon) = \frac{1}{2}\mathbf{L}\varepsilon : \varepsilon \quad \text{and} \quad \mathbf{g}^*(\varepsilon^*) = \frac{1}{2}\mathbf{L}^{-\mathbf{1}}\varepsilon^* : \varepsilon^*.$$

In linear elasticity problems we arrive at the energy functional in terms of strains $\varepsilon(\mathbf{v}) = \frac{1}{2}(\nabla \mathbf{v} + (\nabla \mathbf{v})^{\mathsf{T}})$

$$\frac{1}{2}\int_{\Omega} \mathbb{L}\varepsilon : \varepsilon \,\mathbf{dx} \quad \varepsilon \in \mathbf{L}^2(\Omega, \mathbb{M}^{\mathbf{n}\times\mathbf{n}}),$$

which is mutually dual to the "complementary energy" functional written in terms of stresses $\varepsilon^*(x) => \boldsymbol{\tau}(x)$

$$\frac{1}{2}\int_{\Omega} \mathbb{L}^{-1}\boldsymbol{\tau} : \boldsymbol{\tau} \,\mathbf{dx} \quad \boldsymbol{\tau} \in \mathbf{L}^2(\Omega, \mathbb{M}^{\mathbf{n}\times\mathbf{n}})$$

Example 3 (Nonlinear elasticity, p-Laplacian)

Consider the functional

$$\mathbf{g}(\xi) = \frac{\mathbf{1}}{\mathbf{p}}|\xi|^{\mathbf{p}},$$

where $\mathbf{p} > \mathbf{1}$ and $|\xi| = (\xi \cdot \xi)^{1/2}$. It is easy to verify that the quantity $\xi^* \cdot \xi - \frac{\mathbf{1}}{\mathbf{p}}|\xi|^{\mathbf{p}}$ attains a supremum if $\xi = \xi_0$, where ξ_0 satisfies the relation

$$\xi^* - |\xi_{\mathbf{0}}|^{\mathbf{p}-\mathbf{2}}\xi_{\mathbf{0}} = \mathbf{0},$$

which yields $|\xi^*| = |\xi_0|^{\mathbf{p}-\mathbf{1}}$ and $\xi^* \cdot \xi_0 = |\xi_0|^{\mathbf{p}}$. Therefore,

$$\mathbf{g}^*(\xi^*) = \xi^* \cdot \xi_{\mathbf{0}} - \frac{\mathbf{1}}{\mathbf{p}}|\xi_{\mathbf{0}}|^{\mathbf{p}} = \left(\mathbf{1} - \frac{\mathbf{1}}{\mathbf{p}}\right)|\xi_{\mathbf{0}}|^{\mathbf{p}} = \frac{\mathbf{1}}{\mathbf{p}^*}|\xi^*|^{\mathbf{p}^*},$$

where $\mathbf{p}^* = \frac{\mathbf{p}}{\mathbf{p}-\mathbf{1}}$.

Thus, we obtain another pair of mutually conjugate functionals

$$\mathbf{g}(\boldsymbol{\xi}) = \frac{\mathbf{1}}{\mathbf{p}}|\boldsymbol{\xi}|^{\mathbf{p}} \quad \text{and} \quad \mathbf{g}^*(\boldsymbol{\xi}^*) = \frac{\mathbf{1}}{\mathbf{p}^*}|\boldsymbol{\xi}^*|^{\mathbf{p}^*},$$

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 $\mathbf{y} = \varepsilon(\mathbf{v})$)

$$\frac{1}{\mathbf{p}} \int_{\Omega} |\mathbf{y}|^{\mathbf{p}}\, \mathbf{dx} \quad \mathbf{y} \in \mathbf{L}^{\mathbf{p}}(\Omega, \mathbb{R}^{\mathbf{n}}[\mathbb{M}^{\mathbf{n}\times\mathbf{n}}]),$$

which is mutually dual to

$$\frac{1}{\mathbf{p}^*} \int_{\Omega} |\mathbf{y}^*|^{\mathbf{p}^*}\, \mathbf{dx} \quad \mathbf{y}^* \in \mathbf{L}^{\mathbf{p}^*}(\Omega, \mathbb{R}^{\mathbf{n}}[\mathbb{M}^{\mathbf{n}\times\mathbf{n}}]).$$

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, friction)

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,

$$\mathbf{g}(\boldsymbol{\xi}) = |\boldsymbol{\xi}|, \ \mathbf{g}^*(\boldsymbol{\xi}^*) = \mathfrak{X}_{\mathcal{B}^*(\mathbf{0},\mathbf{1})}(\boldsymbol{\xi}^*), \text{ where } \mathcal{B}^*(\mathbf{0},\mathbf{1}) = \{\boldsymbol{\xi}^* \in \mathbf{E}^{\mathbf{d}} \mid |\boldsymbol{\xi}^*| \leq \mathbf{1}\}$$

Thus, for the functional $\mathbf{G} : \mathbf{L^1} \longrightarrow \mathbb{R}$

$$\mathbf{G(v)} := \int_{\Omega} |\mathbf{v}|\, \mathbf{dx},$$

the respective dual functional is $\mathbf{G}^* : \mathbf{L}^\infty \longrightarrow \mathbb{R}$

$$\mathbf{G}^*(\mathbf{v}^*) = \mathbf{0} \text{ if } |\mathbf{v}^*(\mathbf{x})| \leq \mathbf{1} \text{ a.e. in } \mathbf{\Omega}, \quad \mathbf{G}^*(\mathbf{v}^*) = +\infty \text{ in other cases.}$$

Example 6 (Variational inequalities, perfect plasticity)

Let $\mathbf{K}$ be a convex closed set in $\mathbf{E^d}$ and

$$\mathbf{g}(\boldsymbol{\xi}) = \mathfrak{X}_{\mathbf{K}}(\boldsymbol{\xi}).$$

The respective conjugate functional is defined as follows:

$$\mathbf{g}^*(\boldsymbol{\xi}^*) = \sup_{\boldsymbol{\xi}\in\mathbf{E^d}} \{\boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - \mathfrak{X}_{\mathbf{K}}(\boldsymbol{\xi})\} = \sup_{\boldsymbol{\xi}\in\mathbf{K}} \boldsymbol{\xi}^* \cdot \boldsymbol{\xi}.$$

This function is called the **support** function of $\mathbf{K}$ and is denoted by $\mathfrak{X}^*_{\mathbf{K}}(\boldsymbol{\xi}^*)$. For example, if $\mathbf{K} = \mathcal{B}(\mathbf{0}, \mathbf{1})$, then

$$\sup_{\boldsymbol{\xi}\in\mathbf{K}} \boldsymbol{\xi}^* \cdot \boldsymbol{\xi} = |\boldsymbol{\xi}^*| \;\Rightarrow\; \mathfrak{X}^*_{\mathcal{B}(\mathbf{0},\mathbf{1})}(\boldsymbol{\xi}^*) = |\boldsymbol{\xi}^*|.$$

Example 7 (Elasto-plasticity)

Let us find conjugate for the functional

$$\mathbf{g}^*(\boldsymbol{\xi}^*) = \frac{\mathbf{k}}{\mathbf{2}}|\boldsymbol{\xi}^*|^{\mathbf{2}} + \mathfrak{X}_{\mathcal{B}^*(\mathbf{0},\lambda)}(\boldsymbol{\xi}^*), \quad \mathbf{k} > \mathbf{0},\ \lambda > \mathbf{0}.$$

In this case,

$$\mathbf{g}(\boldsymbol{\xi}) = \sup_{\boldsymbol{\xi}^* \in \mathcal{B}^*(\mathbf{0},\lambda)} \{\boldsymbol{\xi}^* \cdot \boldsymbol{\xi} - \frac{\mathbf{k}}{\mathbf{2}}|\boldsymbol{\xi}^*|^{\mathbf{2}}\}.$$

If $\boldsymbol{\xi}_0^*$ meets the relation $\boldsymbol{\xi} = \mathbf{k}\boldsymbol{\xi}_\mathbf{0}^*$ and satisfies the condition $|\boldsymbol{\xi}_0^*| \leq \lambda$, then it is the required maximizer. For such a $\boldsymbol{\xi}_0^*$ we have

$$\boldsymbol{\xi} \cdot \boldsymbol{\xi}_\mathbf{0}^* - \frac{\mathbf{k}}{\mathbf{2}}|\boldsymbol{\xi}_\mathbf{0}^*|^{\mathbf{2}} = \frac{\mathbf{1}}{\mathbf{k}}|\boldsymbol{\xi}_\mathbf{0}^*|^{\mathbf{2}} - \frac{\mathbf{1}}{\mathbf{2k}}|\boldsymbol{\xi}_\mathbf{0}^*|^{\mathbf{2}} = \frac{\mathbf{1}}{\mathbf{2k}}|\boldsymbol{\xi}|^{\mathbf{2}}.$$

If $|\xi_0^*| > \lambda$, then the maximizer ξ_m^* meets the conditions

$$|\xi_m^*| = \lambda, \quad \xi \cdot \xi_m^* \geq \xi^* \cdot \xi^*, \quad \forall \xi^* \in \mathcal{B}^*(0,\lambda),$$

which mean that $\xi_m^* = \lambda \frac{\xi}{|\xi|}$ and, consequently,

$$\xi \cdot \xi_{\mathbf{m}}^* - \frac{\mathbf{k}}{2}|\xi_{\mathbf{m}}^*|^2 = \lambda|\xi| - \frac{\mathbf{k}}{2}\lambda^2.$$

Thus, we obtain

$$\mathbf{g}(\xi) = \begin{cases} \dfrac{1}{2\mathbf{k}}|\xi|^2 & \text{if } |\xi| \leq \mathbf{k}\lambda, \\ \lambda|\xi| - \dfrac{\mathbf{k}}{2}\lambda^2 & \text{if } |\xi| > \mathbf{k}\lambda. \end{cases}$$

In the theory of perfect elasto–plasticity stresses are subject to the condition $\boxed{\boldsymbol{\tau} \in \mathbf{K} = \textbf{plastic yield set}}$ and the stress energy functional is defined and finite only on such $\boldsymbol{\tau}$:

$$\mathbf{G}^*(\boldsymbol{\tau}) = \frac{\mathbf{1}}{\mathbf{2}} \int_{\boldsymbol{\Omega}} \mathbb{L}^{-\mathbf{1}}\boldsymbol{\tau} : \boldsymbol{\tau}\, \mathbf{dx} \text{ for } \boldsymbol{\tau} \in \mathbf{K}.$$

The respective dual functional (for strains) is given by a linear growth functional

$$\mathbf{G}(\varepsilon) = \int_{\boldsymbol{\Omega}} \mathbf{g}(\varepsilon)\, \mathbf{dx},$$

where **g** is a linear growth functional of the type given on the previous page.

Example 8 (Minimal surfaces, capillary problems)

Consider the functional $\mathbf{g}(\xi) = \sqrt{\mathbf{1} + |\xi|^2}$, arising in some variational problems having a geometrical meaning (e.g., for the nonparametric minimal surface problem). If $|\xi^*| > 1$, then the value of

$$\sup_{\xi \in \mathbf{E}^d} \left\{ \xi^* \cdot \xi - \sqrt{\mathbf{1} + |\xi|^2} \right\}$$

is infinite. If $|\xi^*| \leq 1$, then the maximizer ξ_0 satisfies the condition

$$\xi^* - \frac{\xi_0}{\sqrt{\mathbf{1} + |\xi_0|^2}} = \mathbf{0}, \ \Rightarrow \ |\xi_0|^2 = \frac{|\xi^*|^2}{\mathbf{1} - |\xi^*|^2}.$$

Therefore, we obtain

$$\mathbf{g}^*(\xi^*) = \begin{cases} -\sqrt{1 - |\xi^*|^2} & \text{if } |\xi^*| \leq 1, \\ +\infty & \text{if } |\xi^*| > 1. \end{cases}$$

Energy functional for the minimal surface problem (with $\mathbf{y} = \nabla \mathbf{v}$)

$$\int_\Omega \sqrt{\mathbf{1} + |\mathbf{y}|^2}\,\mathbf{dx} \quad \mathbf{y} \in \mathbf{L}^1(\Omega, \mathbb{R}^2),$$

which is mutually dual to

$$-\int_\Omega \sqrt{\mathbf{1} - |\mathbf{y}^*|^2}\,\mathbf{dx} \quad |\mathbf{y}^*| \leq \mathbf{1}.$$

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 $\mathbf{G}(\mathbf{v}) = \mathbf{J}(\mathbf{v} - \mathbf{v_0})$. Then

$$\mathbf{G}^*(\mathbf{v}^*) = \mathbf{J}^*(\mathbf{v}^*) + \langle \mathbf{v}^*, \mathbf{v_0} \rangle.$$

Proof. Since

$$\sup_{\mathbf{v} \in \mathbf{V}} \{ \langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}(\mathbf{v} - \mathbf{v_0}) \} = \sup_{\mathbf{w} \in \mathbf{V}} \{ \langle \mathbf{v}^*, \mathbf{w} + \mathbf{v_0} \rangle - \mathbf{J}(\mathbf{w}) \}$$
$$= \sup_{\mathbf{w} \in \mathbf{V}} \{ \langle \mathbf{v}^*, \mathbf{w} \rangle - \mathbf{J}(\mathbf{w}) \} + \langle \mathbf{v}^*, \mathbf{v_0} \rangle = \mathbf{J}^*(\mathbf{v}^*) + \langle \mathbf{v}^*, \mathbf{v_0} \rangle,$$

we arrive at the required relation.

Property 5

If $\mathbf{G}(\mathbf{v}) = \min_{\mathbf{i}=\mathbf{1},\dots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}$, then $\mathbf{G}^*(\mathbf{v}^*) = \max_{\mathbf{i}=\mathbf{1},\dots,\mathbf{N}}\{\mathbf{J_i^*}(\mathbf{v}^*)\}$.

Proof. We have

$$\begin{aligned}
\mathbf{G}^*(\mathbf{v}^*) &= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \min_{\mathbf{i}=\mathbf{1},\dots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}\} \\
&= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle + \max_{\mathbf{i}=\mathbf{1},\dots,\mathbf{N}}\{-\mathbf{J_i}(\mathbf{v})\}\} \\
&= \sup_{\mathbf{v}\in\mathbf{V}}\max_{\mathbf{i}=\mathbf{1},\dots,\mathbf{N}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\} \\
&= \max_{\mathbf{i}=\mathbf{1},\dots,\mathbf{N}}\sup_{\mathbf{v}\in\mathbf{V}}\{\langle \mathbf{v}^*, \mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\} = \max_{\mathbf{i}=\mathbf{1},\dots,\mathbf{N}}\{\mathbf{J_i^*}(\mathbf{v}^*)\}.
\end{aligned}$$

Property 6

If $\mathbf{G}(\mathbf{v}) = \max_{\mathbf{i}=1,\ldots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}$, then $\mathbf{G}^*(\mathbf{v}^*) \le \min_{\mathbf{i}=1,\ldots,\mathbf{N}}\{\mathbf{J}^*_\mathbf{i}(\mathbf{v}^*)\}$.

Proof. By definition, we have

$$\begin{aligned}\mathbf{G}^*(\mathbf{v}^*) &= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle\mathbf{v}^*,\mathbf{v}\rangle - \max_{\mathbf{i}=1,\ldots,\mathbf{N}}\{\mathbf{J_i}(\mathbf{v})\}\} \\ &= \sup_{\mathbf{v}\in\mathbf{V}}\{\langle\mathbf{v}^*,\mathbf{v}\rangle + \min_{\mathbf{i}=1,\ldots,\mathbf{N}}\{-\mathbf{J_i}(\mathbf{v})\}\} \\ &= \sup_{\mathbf{v}\in\mathbf{V}}\min_{\mathbf{i}=1,\ldots,\mathbf{N}}\{\langle\mathbf{v}^*,\mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\}\}.\end{aligned}$$

Now we apply $\sup\inf \le \inf\sup$ relation to $\langle\mathbf{v}^*,\mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})$. Then,

$$\mathbf{G}^*(\mathbf{v}^*) \le \min_{\mathbf{i}=1,\ldots,\mathbf{N}}\sup_{\mathbf{v}\in\mathbf{V}}\{\langle\mathbf{v}^*,\mathbf{v}\rangle - \mathbf{J_i}(\mathbf{v})\} = \min_{\mathbf{i}=1,\ldots,\mathbf{N}}\{\mathbf{J}^*_\mathbf{i}(\mathbf{v}^*)\}.$$

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

$$\langle \mathbf{v}^*, \mathbf{v} \rangle - \boldsymbol{\alpha} \leq \mathbf{J}(\mathbf{v}), \quad \forall \mathbf{v} \in \mathbf{V},$$
$$\langle \mathbf{v}^*, \mathbf{v_0} \rangle - \boldsymbol{\alpha} = \mathbf{J}(\mathbf{v_0}).$$

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 (4.1) 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:

$$\mathbf{J}(\mathbf{v}) + \mathbf{J}^*(\mathbf{v}^*) - \langle \mathbf{v}^*, \mathbf{v} \rangle = \mathbf{0}, \tag{4.1}$$

$$\mathbf{v}^* \in \partial \mathbf{J}(\mathbf{v}) \quad \text{and} \quad \mathbf{v} \in \partial \mathbf{J}^*(\mathbf{v}^*). \tag{4.2}$$

Relations (4.2) 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,

$$\mathbf{J}(\mathbf{w}) \geq \mathbf{J}(\mathbf{v}) + \langle \mathbf{v}^*, \mathbf{w} - \mathbf{v} \rangle, \quad \forall \mathbf{w} \in \mathbf{V}.$$

Hence,

$$\langle \mathbf{v}^*, \mathbf{v} \rangle - \mathbf{J}(\mathbf{v}) \geq \langle \mathbf{v}^*, \mathbf{w} \rangle - \mathbf{J}(\mathbf{w}), \quad \forall \mathbf{w} \in \mathbf{V}$$

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:

$$\mathbf{J}(\mathbf{w}) - \mathbf{J}(\mathbf{v}) \geq \langle \mathbf{v}^*, \mathbf{w} - \mathbf{v}\rangle, \quad \forall \mathbf{w} \in \mathbf{V},$$

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, $\mathbf{D_G}(\mathbf{y},\mathbf{y}^*)$ is convex with respect to $\mathbf{y}$ and $\mathbf{y}^*$, but, in general, $\mathbf{D_G}(\mathbf{y},\mathbf{y}^*)$ 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

$$\mathbf{G}(\mathbf{y})=\frac{1}{\boldsymbol{\alpha}}|\mathbf{y}|^{\boldsymbol{\alpha}}\qquad \mathbf{G}^*(\mathbf{y})=\frac{1}{\boldsymbol{\alpha}^*}|\mathbf{y}|^{\boldsymbol{\alpha}^*}.$$

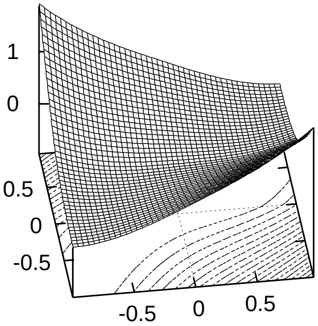
Only for $\boldsymbol{\alpha}=2$ we have a convex functional

$$\mathbf{D_G}(\mathbf{y},\mathbf{y}^*)=\frac{1}{2}|\mathbf{y}|^2+\frac{1}{2}|\mathbf{y}^*|^2-\mathbf{y}\mathbf{y}^*=\frac{1}{2}(\mathbf{y}-\mathbf{y}^*)^2.$$

For other $\boldsymbol{\alpha}\in(1,+\infty)$ $\mathbf{D_G}$ is nonconvex on $\mathbb{R}\times\mathbb{R}$.

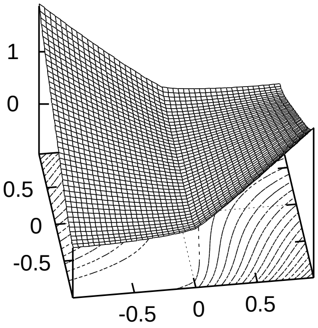
Example 1: $D(\xi_1,\xi_2) = \frac{1}{3}|\xi_1|^3 + \frac{2}{3}|\xi_2|^{3/2} - \xi_1\xi_2$

Compound functional on $\mathbb{R} \times \mathbb{R}$ and its level lines



Example 2: $D(\xi_1, \xi_2) = \frac{5}{6}|\xi_1|^{6/5} + \frac{1}{6}|\xi_2|^6 - \xi_1\xi_2$

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 $\mathbf{y_1^*}, \mathbf{y_2^*} \in \mathbf{Y^*}$,

$$\mathbf{D_G}\left(\frac{\mathbf{y_1}+\mathbf{y_2}}{2}, \frac{\mathbf{y_1^*}+\mathbf{y_2^*}}{2}\right) \leq \frac{1}{4}\Big(\mathbf{D_G}(\mathbf{y_1}, \mathbf{y_1^*}) + \mathbf{D_G}(\mathbf{y_1}, \mathbf{y_2^*}) + \\ + \mathbf{D_G}(\mathbf{y_2}, \mathbf{y_1^*}) + \mathbf{D_G}(\mathbf{y_2}, \mathbf{y_2^*})\Big)$$

Proof

From the definition it follows that

$$\begin{aligned} \mathbf{D_G}\left(\mathbf{y}, \tfrac{\mathbf{y}_1^*+\mathbf{y}_2^*}{2}\right) = \ & \mathbf{G}(\mathbf{y}) + \mathbf{G}^*\left(\tfrac{\mathbf{y}_1^*+\mathbf{y}_2^*}{2}\right) - \left\langle \tfrac{\mathbf{y}_1^*+\mathbf{y}_2^*}{2}, \mathbf{y} \right\rangle \\ & \leq \tfrac{1}{2}\left(\mathbf{D_G}(\mathbf{y}, \mathbf{y}_1^*) + \mathbf{D_G}(\mathbf{y}, \mathbf{y}_2^*)\right) \end{aligned}$$

and

$$\begin{aligned} \mathbf{D_G}\left(\tfrac{\mathbf{y}_1+\mathbf{y}_2}{2}, \mathbf{y}^*\right) = \ & \mathbf{G}\left(\tfrac{\mathbf{y}_1+\mathbf{y}_2}{2}\right) + \mathbf{G}^*(\mathbf{y}^*) - \left\langle \mathbf{y}^*, \tfrac{\mathbf{y}_1+\mathbf{y}_2}{2} \right\rangle \\ & \leq \tfrac{1}{2}\left(\mathbf{D_G}(\mathbf{y}_1, \mathbf{y}^*) + \mathbf{D_G}(\mathbf{y}_2, \mathbf{y}^*)\right). \end{aligned}$$

Therefore,

$$\mathbf{D_G}\left(\tfrac{\mathbf{y}_1+\mathbf{y}_2}{2}, \tfrac{\mathbf{y}_1^*+\mathbf{y}_2^*}{2}\right) \leq \tfrac{1}{2}\left(\mathbf{D_G}\left(\mathbf{y}_1, \tfrac{\mathbf{y}_1^*+\mathbf{y}_2^*}{2}\right) + \mathbf{D_G}\left(\mathbf{y}_2, \tfrac{\mathbf{y}_1^*+\mathbf{y}_2^*}{2}\right)\right).$$

and we arrive at the required estimate.

Important property

If $\mathbf{G}$ and $\mathbf{G}^*$ are Gateaux differentiable, then

$$\langle \mathbf{y}^* - \mathbf{G}'(\mathbf{y}), \mathbf{G}^{*\prime}(\mathbf{y}^*) - \mathbf{y}\rangle \ \geq \ \mathbf{D}_{\mathbf{G}}(\mathbf{y}, \mathbf{y}^*).$$

Note, that from this relation we conclude that $\mathbf{D_J}$ 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 $\mathbf{y_1}, \mathbf{y_2} \in \mathcal{B}(\mathbf{0_Y}, \delta)$ the following inequality holds:

$$\mathbf{J}\left(\frac{\mathbf{y_1} + \mathbf{y_2}}{\mathbf{2}}\right) + \boldsymbol{\Upsilon}_\delta(\mathbf{y_1} - \mathbf{y_2}) \leq \frac{\mathbf{1}}{\mathbf{2}}\left(\mathbf{J}(\mathbf{y_1}) + \mathbf{J}(\mathbf{y_2})\right). \qquad (4.3)$$

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:*

$$\mathbf{J}(\mathbf{z}) \geq \mathbf{J}(\mathbf{y}) + \langle \mathbf{J}'(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle + \mathbf{2}\boldsymbol{\Upsilon}_{\delta}(\mathbf{z} - \mathbf{y})$$

and

$$\langle \mathbf{J}'(\mathbf{z}) - \mathbf{J}'(\mathbf{y}), \mathbf{z} - \mathbf{y} \rangle \geq \mathbf{2}\boldsymbol{\Upsilon}_{\delta}(\mathbf{z} - \mathbf{y}) + \mathbf{2}\boldsymbol{\Upsilon}_{\delta}(\mathbf{y} - \mathbf{z}).$$

Proof.

We have $\Upsilon_\delta(\mathbf{z}-\mathbf{y}) \le \frac{1}{2}\mathbf{J}(\mathbf{z}) + \frac{1}{2}\mathbf{J}(\mathbf{y}) - \mathbf{J}\left(\frac{\mathbf{z}+\mathbf{y}}{2}\right)$.
Since $\mathbf{J}$ is convex and differentiable

$$\mathbf{J}\left(\frac{\mathbf{z}+\mathbf{y}}{2}\right) = \mathbf{J}\left(\mathbf{y} + \frac{\mathbf{z}-\mathbf{y}}{2}\right) \ge \mathbf{J}(\mathbf{y}) + \left\langle \mathbf{J}'(\mathbf{y}), \frac{\mathbf{z}-\mathbf{y}}{2}\right\rangle,$$

and, therefore,

$$2\Upsilon_\delta(\mathbf{z}-\mathbf{y}) \le \mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y}) - \left\langle \mathbf{J}'(\mathbf{y}), \mathbf{z}-\mathbf{y}\right\rangle.$$

We can rewrite it replacing $\mathbf{z}$ by $\mathbf{y}$

$$2\Upsilon_\delta(\mathbf{y}-\mathbf{z}) \le \mathbf{J}(\mathbf{y}) - \mathbf{J}(\mathbf{z}) + \left\langle \mathbf{J}'(\mathbf{z}), \mathbf{z}-\mathbf{y}\right\rangle$$

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* $\mathbf{J}$.

$$\Upsilon_\delta(\mathbf{z} - \mathbf{y_m}) \leq \frac{\mathbf{1}}{\mathbf{2}} (\mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y_m}))\,, \quad \forall \mathbf{z} \in \mathcal{B}(\mathbf{0_Y}, \delta). \tag{4.4}$$

Proof.

Since $\mathbf{J}\left(\frac{\mathbf{y_m}+\mathbf{z}}{\mathbf{2}}\right) \geq \mathbf{J}(\mathbf{y_m})$, we obtain

$$\Upsilon_\delta(\mathbf{z} - \mathbf{y_m}) \leq \quad \frac{1}{2}\mathbf{J}(\mathbf{y_m}) + \frac{\mathbf{1}}{\mathbf{2}}\mathbf{J}(\mathbf{z}) - \mathbf{J}\left(\frac{\mathbf{y_m} + \mathbf{z}}{\mathbf{2}}\right) \leq$$

$$\leq \frac{1}{2} (\mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y_m}))\,.$$

Estimate (4.4) 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 $\boldsymbol{\Upsilon}_\delta$) are controlled by the difference of the functionals.

Corollary 1

Rewrite (4.3) in the form

$$\Upsilon_\delta(\mathbf{z} - \mathbf{y_m}) + \mathbf{J}\left(\frac{\mathbf{y_m} + \mathbf{z}}{2}\right) - \mathbf{J}(\mathbf{y_m}) \leq \frac{1}{2}\left(\mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y_m})\right).$$

By virtue of (4.4), we have

$$\mathbf{J}\left(\frac{\mathbf{y_m} + \mathbf{z}}{2}\right) - \mathbf{J}(\mathbf{y_m}) \geq 2\Upsilon_\delta\left(\frac{\mathbf{z} - \mathbf{y_m}}{2}\right)$$

and, therefore, we arrive at the strengthened estimate

$$\Upsilon_\delta(\mathbf{z} - \mathbf{y_m}) + 2\Upsilon_\delta\left(\frac{\mathbf{z} - \mathbf{y_m}}{2}\right) \leq \frac{1}{2}\left(\mathbf{J}(\mathbf{z}) - \mathbf{J}(\mathbf{y_m})\right). \quad (4.5)$$

Corollary 2

Assume that $\mathbf{J}$ is twice differentiable in the vicinity of $\mathbf{y_m}$ and satisfies the finite increment relation

$$\mathbf{J}\left(\frac{\mathbf{y_m}+\mathbf{z}}{\mathbf{2}}\right) = \mathbf{J}(\mathbf{y_m}) + \left\langle \mathbf{J}'(\mathbf{y_m}), \frac{\mathbf{z}-\mathbf{y_m}}{\mathbf{2}} \right\rangle + \\ +\frac{1}{2}\left\langle \mathbf{J}''\left(\mathbf{y_m}+\xi\frac{\mathbf{z}+\mathbf{y_m}}{\mathbf{2}}\right)\frac{\mathbf{z}-\mathbf{y_m}}{\mathbf{2}}, \frac{\mathbf{z}-\mathbf{y_m}}{\mathbf{2}} \right\rangle,$$

where $\boldsymbol{\xi} \in (0,1)$. Since $\mathbf{J}'(\mathbf{y_m}) = \mathbf{0_{Y^*}}$, we have another estimate:

$$\boldsymbol{\Upsilon}_\delta(\mathbf{z}-\mathbf{y_m}) + \frac{\mathbf{1}}{\mathbf{8}}\left\langle \mathbf{J}''\left((\mathbf{1}+\tfrac{\boldsymbol{\xi}}{\mathbf{2}})\mathbf{y_m}+\tfrac{\boldsymbol{\xi}}{\mathbf{2}}\mathbf{z}\right)(\mathbf{z}-\mathbf{y_m}), \mathbf{z}-\mathbf{y_m} \right\rangle \leq \\ \leq \frac{1}{2}\left(\mathbf{J}(\mathbf{z})-\mathbf{J}(\mathbf{y_m})\right). \quad (4.6)$$

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 (4.4) we have

$$\frac{1}{8}(\mathbf{A}(\mathbf{z}-\mathbf{y_m}),(\mathbf{z}-\mathbf{y_m})) \leq \frac{1}{2}\left(\mathbf{J}(\mathbf{z})-\mathbf{J}(\mathbf{y_m})\right), \quad \forall \mathbf{z}$$

However (4.6) gives a better estimate

$$\frac{1}{2}(\mathbf{A}(\mathbf{z}-\mathbf{y_m}),(\mathbf{z}-\mathbf{y_m})) \leq \mathbf{J}(\mathbf{z})-\mathbf{J}(\mathbf{y_m}). \tag{4.7}$$

Note that for quadratic type functionals this estimate holds as equality. Indeed,

$$\mathbf{J}\left(\mathbf{z}\right)-\mathbf{J}(\mathbf{y_m})=(\mathbf{A}\mathbf{y_m}+\ell,\mathbf{z}-\mathbf{y_m})+\frac{1}{2}(\mathbf{A}(\mathbf{z}-\mathbf{y_m}),\mathbf{z}-\mathbf{y_m}).$$

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, (4.7) **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

$$\mathbf{J}(\mathbf{y}) = \frac{1}{2}(\mathbf{A}\mathbf{y}, \mathbf{y}) + (\ell, \mathbf{y}) + \boldsymbol{\Psi}(\mathbf{y}),$$

where $\Psi(y)$ is a convex and l.s.c. functional. Applying the above Theorem with $\mu_1 = \mu_2 = 1$,

$$\mathbf{J_1}(\mathbf{y}) = \frac{1}{2}(\mathbf{A}\mathbf{y}, \mathbf{y}) + (\ell, \mathbf{y}) \qquad \mathbf{J_2}(\mathbf{y}) = \boldsymbol{\Psi}(\mathbf{y}),$$

we see that $\mathbf{J}$ is uniformly convex with functional $\boldsymbol{\Upsilon}$ 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* $\boldsymbol{\Upsilon}_{\mathbf{1}\delta}$ *and* $\boldsymbol{\Upsilon}_{\mathbf{2}\delta}$*, 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*

$$\boldsymbol{\Upsilon}_{\delta} = \min\{\boldsymbol{\Upsilon}_{\mathbf{1}\delta}, \boldsymbol{\Upsilon}_{\mathbf{2}\delta}\}.$$

Proof. We have

$$\frac{1}{2}\mathbf{J}(\mathbf{y}) + \frac{1}{2}\mathbf{J}(\mathbf{z}) - \mathbf{J}\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right) = \frac{1}{2}\max\{\mathbf{J}_1(\mathbf{y}), \mathbf{J}_2(\mathbf{y})\} + \\ + \frac{1}{2}\max\{\mathbf{J}_1(\mathbf{z}), \mathbf{J}_2(\mathbf{z})\} - \max\left\{\mathbf{J}_1\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right), \mathbf{J}_2\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right)\right\}.$$

Assume that

$$\max\left\{\mathbf{J}_1\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right), \mathbf{J}_2\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right)\right\} = \mathbf{J}_1\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right).$$

Then

$$\frac{1}{2}\left(\mathbf{J}(\mathbf{y}) + \mathbf{J}(\mathbf{z})\right) - \mathbf{J}\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right) \geq \\ \geq \frac{1}{2}\left(\mathbf{J}_1(\mathbf{y}) + \mathbf{J}_1(\mathbf{z})\right) - \mathbf{J}_1\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right) \geq \boldsymbol{\Upsilon}_{1\delta}(\mathbf{z} - \mathbf{y}).$$

If we have an opposite situation, i.e.,

$$\max\left\{\mathbf{J_1}\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right), \mathbf{J_2}\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right)\right\} = \mathbf{J_2}\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right),$$

then

$$\frac{1}{2}\mathbf{J}(\mathbf{y}) + \frac{\mathbf{1}}{\mathbf{2}}\mathbf{J}(\mathbf{z}) - \mathbf{J}\left(\frac{\mathbf{y}+\mathbf{z}}{2}\right) \geq \boldsymbol{\Upsilon}_{\mathbf{2}\delta}(\mathbf{z}-\mathbf{y}).$$

Thus, in both cases the lower bound is given by the functional

$$\min\ \{\boldsymbol{\Upsilon}_{\mathbf{1}\delta}(\mathbf{z}\ -\ \mathbf{y}), \boldsymbol{\Upsilon}_{\mathbf{2}\delta}(\mathbf{z}\ -\ \mathbf{y})\}.$$

Example 3. Power growth functionals

Let

$$\mathbf{G}(\mathbf{y}) = \frac{1}{\alpha}\int_{\Omega} |\mathbf{y}|^{\alpha}\, \mathbf{dx} \qquad \mathbf{F}(\mathbf{v}) = \int_{\Omega} \mathbf{fv}\mathbf{dx},$$

where $\alpha > 1$. Then Problem $\mathcal{P}$ is to minimize the functional

$$\mathbf{J}_{\alpha}(\mathbf{v}) := \int_{\Omega} \left(\frac{1}{\alpha}|\nabla \mathbf{v}|^{\alpha} + \mathbf{fv}\right) \mathbf{dx}$$

over the space $\mathbf{V} = \{\mathbf{v} \in \mathbf{H}^{\alpha}(\Omega) \mid \mathbf{v} = \mathbf{0} \text{ on } \partial\Omega\}$.

Problem $\mathcal{P}^*$ is to maximize the functional

$$\mathbf{I}^*_{\boldsymbol{\alpha}_*}(\mathbf{y}^*) = -\frac{1}{\boldsymbol{\alpha}_*}\int_{\boldsymbol{\Omega}} |\mathbf{y}^*|^{\boldsymbol{\alpha}^*}\,\mathbf{dx}$$

over the set

$$\mathbf{Q}^*_{\mathbf{f}} = \left\{ \mathbf{y}^* \in \mathbf{Y}^* := \mathbf{L}^{\boldsymbol{\alpha}_*}(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}}) | \int_{\boldsymbol{\Omega}} \mathbf{y}^* \cdot \nabla \mathbf{w} \mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{f} \mathbf{w} \mathbf{dx} \ \forall \mathbf{w} \in \mathbf{V} \right\}.$$

For $\alpha \geq 2$ uniform convexity of $\mathbf{G}(\mathbf{y})$ follows from the first Clarkson's inequality

$$\int_\Omega \left|\frac{\mathbf{y_1}+\mathbf{y_2}}{2}\right|^\alpha \mathbf{dx} + \int_\Omega \left|\frac{\mathbf{y_1}-\mathbf{y_2}}{2}\right|^\alpha \mathbf{dx} \leq \frac{1}{2}\int_\Omega \left(|\mathbf{y_1}|^\alpha + |\mathbf{y_2}|^\alpha\right)\mathbf{dx},$$

which is valid for all $\mathbf{y_1},\ \mathbf{y_2} \in \mathbf{Y}$.

See S. L. Sobolev. *Some Applications of Functional Analysis in Mathematical Physics*. Hence, we observe that in this case

$$\Upsilon_\ominus(\mathbf{z}) = \frac{1}{\alpha}\,\|\mathbf{z}\|^\alpha_{\alpha,\Omega}\,.$$

and

$$\frac{1}{\alpha 2^\alpha}\int_\Omega |\nabla(\mathbf{v}-\mathbf{u})|^\alpha \mathbf{dx} \leq \frac{1}{2}\left(\mathbf{J}_\alpha(\mathbf{v}) - \mathbf{I}^*_\alpha(\mathbf{q}^*)\right),\ \forall \mathbf{q}^* \in \mathbf{Q}^*_\mathbf{f},$$

For $1 < \alpha \leq 2$, the functional $\mathbf{G}$ is also uniformly convex. This fact follows from the second Clarkson's inequality

$$\left(\int_{\Omega}\left(\frac{y_1+y_2}{2}\right)^{\alpha}dx\right)^{\frac{1}{\alpha-1}}+\left(\int_{\Omega}\left(\frac{y_1-y_2}{2}\right)^{\alpha}dx\right)^{\frac{1}{\alpha-1}} \leq \left(\frac{1}{2}\int_{\Omega}\left(|y_1|^{\alpha}+|y_2|^{\alpha}\right)dx\right)^{\frac{1}{\alpha-1}}.$$

However, in this case, the functional Υ_δ 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

$$\Upsilon_\delta(\mathbf{z}) = \delta^{\frac{\alpha-2}{\alpha-1}}\kappa\,\|\mathbf{z}\|_{\alpha,\Omega}^{\frac{\alpha}{\alpha-1}},$$

where $\kappa = \frac{1}{\kappa_0+1}$ and κ_0 is the integer part of $\frac{1}{\alpha-1}$.

Lecture 5.
FUNCTIONAL A POSTERIORI ESTIMATES. GENERAL APPROACH.

Main goal of the lecture

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

$$\Lambda^* \mathcal{A} \Lambda \mathbf{u} + \ell = 0$$

where Λ 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;

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**

$$(\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda w}) + \langle \ell, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0}. \tag{5.1}$$

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

$$\mathbf{V} \quad \text{is compactly embedded in} \quad \mathbf{U} \tag{5.2}$$

and the operators $\mathbf{\Lambda}$ and $\mathcal{A}$ satisfy the relations

$$c_1\|\mathbf{y}\|^2 \leq (\mathcal{A}\mathbf{y},\mathbf{y}) \leq c_2\|\mathbf{y}\|^2, \qquad \forall \mathbf{y} \in \mathbf{U}, \tag{5.3}$$

$$\|\mathbf{\Lambda} \mathbf{w}\| \geq c_3\|\mathbf{w}\|_{\mathbf{V}}, \quad \forall \mathbf{w} \in \mathbf{V_0}, \tag{5.4}$$

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 $\mathcal{A}^{-1}$ is the operator inverse to $\mathcal{A}$. The respective spaces Y and Y^* contain elements of $\mathbf{U}$ equipped with the norms $\|\!| \cdot \|\!|$ and $\|\!| \cdot \|\!|_*$, respectively.

Problem (5.1) is equivalent to following problem.

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

$$\mathbf{J}(\mathbf{v}) = \frac{1}{2} \|\!| \boldsymbol{\Lambda}\mathbf{v} \|\!|^2 + \langle \ell, \mathbf{v} \rangle.$$

Lagrangian

On the set $(\mathbf{V_0} + \mathbf{u_0}) \times \mathbf{Y}^*$, we define the Lagrangian

$$\mathbf{L}(\mathbf{v}, \mathbf{y}) = (\mathbf{y}, \mathbf{\Lambda v}) - \frac{1}{2} \mathbin{|\!|\!|} \mathbf{y} \mathbin{|\!|\!|}^2 + \langle \ell, \mathbf{v} \rangle$$

and the functional

$$\mathbf{l}^*(\mathbf{y}) = \inf_{\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}} \mathbf{L}(\mathbf{v}, \mathbf{y}) = \begin{cases} (\mathbf{y}, \mathbf{\Lambda u_0}) - \frac{1}{2} \mathbin{|\!|\!|} \mathbf{y} \mathbin{|\!|\!|}_*^2 + \langle \ell, \mathbf{u_0} \rangle, & \mathbf{y} \in \mathbf{Q}_\ell^*, \\ -\infty, & \mathbf{y} \notin \mathbf{Q}_\ell^*, \end{cases}$$

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 $\mathbf{l}^*$ 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

$$\mathbf{I}^*(\mathbf{p}) = \sup_{\mathbf{y} \in \mathbf{Q}^*_\ell} \mathbf{I}^*(\mathbf{y}) := \sup \mathcal{P}^* \leq \inf \mathcal{P}.$$

The minimizer **u** satisfies and the maximizer **p** satisfies the **stationarity conditions**

$$(\mathcal{A}\Lambda\mathbf{u}, \Lambda\mathbf{w}) + \langle \ell, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0},$$
$$(\Lambda\mathbf{u_0} - \mathcal{A}^{-1}\mathbf{p}, \mathbf{y}) = \mathbf{0}, \quad \forall \mathbf{y} \in \mathbf{Q}^*_\mathbf{0},$$

where $\mathbf{Q}^*_\mathbf{0} := \left\{ \mathbf{y} \in \mathbf{Y}^* \middle| (\mathbf{y}, \Lambda\mathbf{w}) = \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0} \right\}.$

We see that $\mathcal{A}\Lambda\mathbf{u} \in \mathbf{Q}^*_\ell$,.

Take

$$\mathbf{I}^*(\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}) = (\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}, \boldsymbol{\Lambda}\mathbf{u_0}) - \frac{1}{2} \, \| \, \boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u} \, \|_*^2 + \langle \ell, \mathbf{u_0} \rangle$$

and set $\mathbf{u_0} = \mathbf{u}$. We obtain

$$\mathbf{I}^*(\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}) = (\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}, \boldsymbol{\Lambda}\mathbf{u}) - \frac{1}{2} \, \| \, \boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u} \, \|_*^2 + \langle \ell, \mathbf{u} \rangle \leq \sup \mathcal{P}^*.$$

Since $\| \, \boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u} \, \|_*^2 = (\boldsymbol{\mathcal{A}}^{-1}\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}, \boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}) = \| \, \boldsymbol{\Lambda}\mathbf{u} \, \|^2$, we see that

$$\mathbf{I}^*(\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}) = \mathbf{J}(\mathbf{u}) = \inf \mathcal{P}$$

Thus

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

The relation $\mathbf{I}^*(\mathbf{p}) = \mathbf{J}(\mathbf{u})$ means that

$$(\mathbf{p}, \mathbf{\Lambda u}) - \frac{1}{2} \parallel \mathbf{p} \parallel_*^2 + \langle \ell, \mathbf{u} \rangle = \frac{1}{2} \parallel \mathbf{\Lambda u} \parallel^2 + \langle \ell, \mathbf{u} \rangle,$$

which is equivalent to the relation

$$\mathbf{D}(\mathbf{\Lambda u}, \mathbf{p}) = \frac{1}{2} \parallel \mathbf{\Lambda u} \parallel^2 + \frac{1}{2} \parallel \mathbf{p} \parallel_*^2 - (\mathbf{p}, \mathbf{\Lambda u}) = \mathbf{0}.$$

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 $v \in \mathbf{V_0} + \mathbf{u_0}$ 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 $\mathbf{q} \in \mathbf{Q}^*_\ell$,

$$\|\, \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \,\|^2 + \|\, \mathbf{q}-\mathbf{p} \,\|_*^2 = 2\,(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q})), \tag{5.5}$$

$$\|\, \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \,\|^2 + \|\, \mathbf{q}-\mathbf{p} \,\|_*^2 = 2\,\mathbf{D}(\mathbf{\Lambda v}, \mathbf{q}). \tag{5.6}$$

Proof

By the stationarity relations, we have

$$\begin{aligned}\tfrac{1}{2}\,|\!|\!|\, \mathbf{\Lambda}(\mathbf{v}-\mathbf{u})\,|\!|\!|^2 &= \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) + \\ &(\mathcal{A}\mathbf{\Lambda}\mathbf{u}, \mathbf{\Lambda}(\mathbf{u}-\mathbf{v})) + \langle \ell, \mathbf{u}-\mathbf{v}\rangle = \\ &= \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}).\end{aligned}$$

Analogously

$$\begin{aligned}\tfrac{1}{2}\,|\!|\!|\, \mathbf{q}-\mathbf{p}\,|\!|\!|_*^2 &= \mathbf{I}^*(\mathbf{p}) - \mathbf{I}^*(\mathbf{q}) + (\mathbf{\Lambda}\mathbf{u_0} - \mathcal{A}^{-1}\mathbf{p}, \mathbf{p}-\mathbf{q}) = \\ &= \mathbf{I}^*(\mathbf{p}) - \mathbf{I}^*(\mathbf{q}).\end{aligned}$$

Since $\mathbf{J}(\mathbf{u}) = \mathbf{I}^*(\mathbf{p})$, we sum two relations and obtain (5.5). 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 (5.6) follows from (5.5).

The estimates (5.5) and (5.6) 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*

$$\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q}) \leq (\mathbf{1}+\beta)\mathbf{D}(\boldsymbol{\Lambda}\mathbf{v}, \mathbf{y}) + \frac{\mathbf{1}+\beta}{\mathbf{2}\beta} \ \| \ \mathbf{q} - \mathbf{y} \ \|_*^{\mathbf{2}} \ . \quad (5.7)$$

Note that

$$\begin{aligned}
\mathbf{D}(\boldsymbol{\Lambda}\mathbf{v}, \mathbf{y}) &= \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}\boldsymbol{\Lambda}\mathbf{v}, \boldsymbol{\Lambda}\mathbf{v}) + \frac{\mathbf{1}}{\mathbf{2}}(\mathcal{A}^{-\mathbf{1}}\mathbf{p}, \mathbf{p}) - (\mathbf{y}, \boldsymbol{\Lambda}\mathbf{u}) = \\
&= (\mathcal{A}\boldsymbol{\Lambda}\mathbf{v} - \mathbf{y}), \boldsymbol{\Lambda}\mathbf{v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y}) = \\
&= (\mathcal{A}(\boldsymbol{\Lambda}\mathbf{v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y}, \boldsymbol{\Lambda}\mathbf{v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y}) = \| \ \boldsymbol{\Lambda}\mathbf{v} - \mathcal{A}^{-\mathbf{1}}\mathbf{y} \ \|.
\end{aligned}$$

Proof

For any $\mathbf{y} \in \mathbf{Y}^*$, we have

$$\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q}) = \frac{1}{2}\Big(\|\!| \Lambda\mathbf{v} |\!\|^2 + \|\!| \mathbf{y} |\!\|_*^2\Big) + \\ + \frac{1}{2}\big(\|\!| \mathbf{q} |\!\|_*^2 - \|\!| \mathbf{y} |\!\|_*^2\big) - (\Lambda\mathbf{u_0}, \mathbf{q}) + \langle \ell, \mathbf{v} - \mathbf{u_0}\rangle.$$

Since $\langle \ell, \mathbf{v} - \mathbf{u_0}\rangle = (\mathbf{q}, \Lambda(\mathbf{u_0} - \mathbf{v}))$, we find that

$$\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q}) = \frac{1}{2}\left(\|\!| \Lambda\mathbf{v} |\!\|^2 + \|\!| \mathbf{y} |\!\|_*^2\right) + \frac{1}{2}\left(\|\!| \mathbf{q} |\!\|_*^2 - \|\!| \mathbf{y} |\!\|_*^2\right) - (\mathbf{q}, \Lambda\mathbf{v}) = \\ = \mathbf{D}(\Lambda\mathbf{v}, \mathbf{y}) + \Big(\mathbf{y} - \mathbf{q}, \Lambda\mathbf{v} - \mathcal{A}^{-1}\mathbf{y}\Big) + \frac{1}{2}\|\!| \mathbf{q} - \mathbf{y} |\!\|_*^2 .$$

This relation yields (5.7) if we use the Young's inequality

$$2\left(\mathbf{y} - \mathbf{q}, \Lambda\mathbf{v} - \mathcal{A}^{-1}\mathbf{y}\right) \le \beta \|\!| \Lambda\mathbf{v} - \mathcal{A}^{-1}\mathbf{y} |\!\|^2 + \beta^{-1} \|\!| \mathbf{y} - \mathbf{q} |\!\|_*^2 .$$

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, (5.7) imply the estimate

$$\frac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 \leq (1+\beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \left(1+\frac{1}{\beta}\right)\frac{1}{2}\mathbf{d}_\ell{}^2(\mathbf{y})$$

where $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$ and $\mathbf{y} \in \mathbf{Y}^*$. We rewrite this estimate as

$$\frac{1}{2} \| \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 \leq \mathcal{M}(\mathbf{v},\beta), \quad \forall \mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}, \quad \beta \in \mathbb{R}_+,$$

where

$$\mathcal{M}(\mathbf{v},\beta) := \inf_{\mathbf{y}\in\mathbf{Y}^*} \left\{ (1+\beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \left(1+\frac{1}{\beta}\right)\frac{1}{2}\mathbf{d}_\ell^2(\mathbf{y}) \right\}.$$

Above estimate is sharp for any $\boldsymbol{\beta}$!

Theorem

For any $\boldsymbol{\beta} \in \mathbb{R}_+$,

$$\frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^{\mathbf{2}} = \boldsymbol{\mathcal{M}}(\mathbf{v}, \boldsymbol{\beta}).$$

Proof. Set $\mathbf{y} = \lambda \mathbf{p} + (\mathbf{1} - \lambda)\boldsymbol{\mathcal{A}}\mathbf{\Lambda v}$. Then $\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) = \frac{\mathbf{1}}{\mathbf{2}}\lambda^{\mathbf{2}} \| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^{\mathbf{2}}$.
Since

$$d_\ell^2(\mathbf{y}) \leq \| \mathbf{p} - \mathbf{y} \|_*^{\mathbf{2}} = \\ = (1-\lambda)^2 \| p - \boldsymbol{\mathcal{A}}\mathbf{\Lambda} v \|_*^2 = (1-\lambda)^2 \| \boldsymbol{\mathcal{A}}\mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|_*^{\mathbf{2}} = \\ = (1-\lambda)^2 \| \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \|^{\mathbf{2}},$$

we obtain

$$\boldsymbol{\mathcal{M}}(\mathbf{v}, \boldsymbol{\beta}) \leq \frac{\mathbf{1}}{\mathbf{2}} \left((\mathbf{1} + \boldsymbol{\beta})\lambda^{\mathbf{2}} + \left(\mathbf{1} + \frac{\mathbf{1}}{\boldsymbol{\beta}}\right) (\mathbf{1} - \lambda)^{\mathbf{2}} \right) \| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^{\mathbf{2}} .$$

The right-hand side attains its minimal value at $\lambda = 1/(1+\beta)$, which leads to the estimate

$$\frac{1}{2} \parallel \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \parallel^2 \geq \mathcal{M}(\mathbf{v},\beta), \quad \forall \mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}, \quad \beta \in \mathbb{R}_+.$$

Recalling that the inverse inequality has already been established, we arrive at the required conclusion

Now, we proceed to finding computable upper bounds for the quantity $\mathbf{d}_\ell$. The first step is given by

Theorem

$$\frac{1}{2}\mathbf{d}_\ell^2(\mathbf{y}) \; = \; \sup_{\mathbf{w} \in \mathbf{V_0}} \left\{ -\frac{1}{2} \parallel \mathbf{\Lambda w} \parallel^2 - \langle \ell, \mathbf{w} \rangle - (\mathbf{y}, \mathbf{\Lambda w}) \right\}.$$

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 $\mathbf{Q}_\ell^* - \mathbf{y} := \left\{ \boldsymbol{\eta}^* \in \mathbf{Y}^* \middle| \boldsymbol{\eta}^* = \text{æ}^* - \mathbf{y}, \quad \text{æ}^* \in \mathbf{Q}_\ell^* \right\}$.
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

$$\frac{1}{2}\mathbf{d}^2_\ell(\mathbf{y}) = - \sup_{\eta^* \in \mathbf{Q}^*_{\ell_\mathbf{y}}} \left\{ -\frac{1}{2} \parallel \eta^* \parallel^2_* \right\}.$$

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

$$\begin{aligned} \frac{1}{2}\mathbf{d}^2_\ell(\mathbf{y}) &= - \inf_{w \in \mathbf{V_0}} \left\{ \frac{1}{2} \parallel \mathbf{\Lambda w} \parallel^2 + \langle \ell_\mathbf{y}, \mathbf{w} \rangle \right\} = \\ &= - \inf_{w \in \mathbf{V_0}} \left\{ \frac{1}{2} \parallel \mathbf{\Lambda w} \parallel^2 + \langle \ell, \mathbf{w} \rangle + (\mathbf{y}, \mathbf{\Lambda w}) \right\} = \\ &= \sup_{w \in \mathbf{V_0}} \left\{ -\frac{1}{2} \parallel \mathbf{\Lambda w} \parallel^2 - \langle \ell, \mathbf{w} \rangle - (\mathbf{y}, \mathbf{\Lambda w}) \right\}. \end{aligned}$$

□

Corollary

We arrive at the conclusion that the majorant $\mathcal{M}(\mathbf{v}, \beta)$ has a minimax form

$$\mathcal{M}(\mathbf{v}, \beta) = \inf_{\mathbf{y} \in \mathbf{Y}^*} \sup_{\mathbf{w} \in \mathbf{V_0}} \left\{ (1+\beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \frac{\mathbf{1}+\beta}{\beta} \Big(-(\mathbf{y}, \mathbf{\Lambda w}) - \mathbf{J}(\mathbf{w}) \Big) \right\}. \quad (5.8)$$

Further,we use (5.8) for deriving upper and lower error bounds.

Upper estimates of $|\!|\!| \mathbf{v} - \mathbf{u} |\!|\!|$

In the relation

$$\mathcal{M}(\mathbf{v}, \beta) \leq (\mathbf{1} + \beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) +$$
$$+ \left(1 + \frac{1}{\beta}\right) \sup_{w \in \mathbf{V_0}} \left\{ -\frac{\mathbf{1}}{\mathbf{2}} |\!|\!| \mathbf{\Lambda w} |\!|\!|^{\mathbf{2}} - \langle \ell, \mathbf{w} \rangle - (\mathbf{y}, \mathbf{\Lambda w}) \right\},$$

we will estimate the value of supremum. Let $\mathbf{\Lambda}^*$ be the operator conjugate to $\mathbf{\Lambda}$, i.e.,

$$(\mathbf{y}, \mathbf{\Lambda w}) = \langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle, \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

Then

$$\langle \ell, w \rangle + \langle \mathbf{y}, \mathbf{\Lambda w} \rangle = \langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle \leq [\![\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,]\!] \, \|\mathbf{\Lambda w}\|.$$

Here

$$[\![\, \ell + \Lambda^* y \,]\!] := \sup_{w \in V_0} \frac{\langle \ell + \Lambda^* y, w \rangle}{|\!|\!| \Lambda w |\!|\!|} < +\infty.$$

To prove that the value of the negative norm is finite we estimate the numerator as follows:

$$\langle \ell + \Lambda^* y, w \rangle \leq \|\ell\|_{V_0^*} \|w\|_V + \|y\| \|\Lambda w\| \leq \left(c_3^{-1} \|\ell\|_{V_0^*} + \|y\| \right) \|\Lambda w\| \leq$$
$$\leq c_1^{-1/2} \left(c_3^{-1} \|\ell\|_{V_0^*} + \|y\| \right) |\!|\!| \Lambda w |\!|\!| \,.$$

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

$$\frac{1}{2} \||\, \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \,\||^2 \leq (\mathbf{1}+\beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \frac{\mathbf{1}+\beta}{\mathbf{2}\beta} [\![\, \ell + \mathbf{\Lambda}^*\mathbf{y} \,]\!]^{\mathbf{2}}. \quad (5.9)$$

This estimate contains the norm $[\![\cdot]\!]$ defined via a sup-relation. We replace it by the norm in a Hilbert space $\mathbf{U}$ provided that ℓ belongs to a narrower set. Assume that

$$\ell \in \mathbf{U} \subset \mathbf{V}_{\mathbf{0}}^*,$$
$$\mathbf{y} \in \mathbf{Q}^* := \{\mathbf{z}^* \in \mathbf{Y}^* | \quad \mathbf{\Lambda}^*\mathbf{z}^* \in \mathbf{U}\}.$$

Note that $\mathbf{Q}^*$ can be endowed with the norm

$$\|\mathbf{y}\|^{\mathbf{2}}_{\mathbf{Q}^*} := \|\mathbf{y}\|^{\mathbf{2}}_* + \|\mathbf{\Lambda}^*\mathbf{y}^*\|^{\mathbf{2}}_{\mathbf{U}}.$$

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.

Majorant of the deviation

Then

$$\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = (\boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w}) \quad \mathbf{w} \in \mathbf{V_0}.$$

$$[\![\, \ell + \boldsymbol{\Lambda}^* \mathbf{y} \,]\!] = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle}{\|\!|\, \boldsymbol{\Lambda} \mathbf{w} \,|\!\|} \leq \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\|\boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}\| \, \|\mathbf{w}\|}{\|\!|\, \boldsymbol{\Lambda} \mathbf{w} \,|\!\|} \leq$$

$$\leq \|\boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}\| \mathbf{c_1}^{-1} \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\|\mathbf{w}\|}{\|\boldsymbol{\Lambda} \mathbf{w}\|} \leq \mathbf{c_1}^{-1} \mathbf{c_3}^{-1} \|\boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}\|.$$

Here $\mathbf{c_1}$ and $\mathbf{c_3}$ are the constants in (5.3) and (5.4). Denote $\mathbf{c}^2 = \mathbf{c_1}^{-2} \mathbf{c_3}^{-2}$. Now, the Majorant is represented in the form

$$\frac{1}{2} \|\!|\, \boldsymbol{\Lambda}(\mathbf{v} - \mathbf{u}) \,|\!\|^2 \leq \mathbf{M}_{\oplus}(\mathbf{v}, \boldsymbol{\beta}, \mathbf{y}) :=$$

$$:= (1 + \boldsymbol{\beta}) \mathbf{D}(\boldsymbol{\Lambda} \mathbf{v}, \mathbf{y}) + \frac{1 + \boldsymbol{\beta}}{2 \boldsymbol{\beta}} \mathbf{c}^2 \|\boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}\|^2. \quad (5.10)$$

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 $v \in \mathbf{V_0} + \mathbf{u_0}$,

$$\frac{1}{2} \mid\!\mid\!\mid \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \mid\!\mid\!\mid^2 = \inf_{\substack{\mathbf{y} \in \mathbf{Q}^* \\ \boldsymbol{\beta} > \mathbf{0}}} \mathbf{M}_\oplus(\mathbf{v}, \beta, \mathbf{y}).$$

If $\ell \in \mathbf{U}$, then $\mathbf{p} \in \mathbf{Q}^*$ and, therefore,

$$\inf_{\substack{\mathbf{y} \in \mathbf{Q}^* \\ \boldsymbol{\beta} > \mathbf{0}}} \mathbf{M}_\oplus(\mathbf{v}, \beta, \mathbf{y}) \leq \mathbf{M}_\oplus(\mathbf{v}, \varepsilon, \mathbf{p}) = (\mathbf{1} + \varepsilon)\frac{\mathbf{1}}{\mathbf{2}} \mid\!\mid\!\mid \mathbf{\Lambda}(\mathbf{u} - \mathbf{v}) \mid\!\mid\!\mid^2,$$

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 $\sup\inf \leq \inf\sup$, we have

$$\mathcal{M}(\mathbf{v},\beta) \geq \sup_{w\in\mathbf{V_0}}\inf_{\mathbf{y}\in\mathbf{Y}^*}\Big\{(1+\beta)D(\mathbf{\Lambda}v,\mathbf{y})-$$

$$-\left(1+\frac{1}{\beta}\right)\left(\frac{\mathbf{1}}{\mathbf{2}}\,\|\!|\,\mathbf{\Lambda w}\,\|\!|^{\mathbf{2}}+\langle\ell,\mathbf{w}\rangle+(\mathbf{y},\mathbf{\Lambda w})\right)\Big\}.$$

Thus, for any $\mathbf{w} \in \mathbf{V_0}$

$$\mathcal{M}(\mathbf{v}, \beta) \geq$$
$$\inf_{\mathbf{y} \in \mathbf{Y}^*} \left\{ (1+\beta) \left(\frac{1}{2} \| \mathbf{y} \|_*^2 - (\mathbf{y}, \Lambda \mathbf{v}) \right) - \left(1 + \frac{1}{\beta} \right) (\mathbf{y}, \Lambda \mathbf{w}) \right\} +$$
$$+ (1+\beta) \frac{1}{2} \| \Lambda \mathbf{v} \|^2 - \left(1 + \frac{1}{\beta} \right) \left(\frac{1}{2} \| \Lambda \mathbf{w} \|^2 + \langle \ell, \mathbf{w} \rangle \right),$$

Evidently, this estimate is also valid for the function $\beta \mathbf{w}$, which yields

$$\mathcal{M}(\mathbf{v}, \beta) \geq (1+\beta) \inf_{\mathbf{y} \in \mathbf{Y}^*} \left\{ \frac{1}{2} \| \mathbf{y} \|_*^2 - (\mathbf{y}, \Lambda(\mathbf{v} + \mathbf{w})) \right\} +$$
$$+ (1+\beta) \left(\frac{1}{2} \| \Lambda \mathbf{v} \|^2 - \frac{\beta}{2} \| \Lambda \mathbf{w} \|^2 - \langle \ell, \mathbf{w} \rangle \right).$$

Note that

$$\inf_{\mathbf{y}\in\mathbf{Y}^*}\left\{\frac{1}{2}\parallel \mathbf{y}\parallel_*^2-(\mathbf{y},\mathbf{\Lambda}(\mathbf{v}+\mathbf{w}))\right\}\geq$$
$$\geq\inf_{\mathbf{y}\in\mathbf{Y}^*}\left\{\frac{1}{2}\parallel \mathbf{y}\parallel_*^2-\parallel \mathbf{y}\parallel_*\parallel \mathbf{\Lambda}(\mathbf{v}+\mathbf{w})\parallel\right\}=-\frac{1}{2}\parallel \mathbf{\Lambda}(\mathbf{v}+\mathbf{w})\parallel^2 .$$

Thus, we obtain

$$\mathcal{M}(\mathbf{v},\beta)\geq(1+\beta)\Big\{-\frac{1}{2}\parallel \mathbf{\Lambda}(\mathbf{v}+\mathbf{w})\parallel^2+$$
$$+\frac{1}{2}\parallel \mathbf{\Lambda v}\parallel^2-\frac{\beta}{2}\parallel \mathbf{\Lambda w}\parallel^2-\langle\ell,\mathbf{w}\rangle\Big\}=$$
$$=(1+\beta)\Big\{-(\mathcal{A}\mathbf{\Lambda}v,\mathbf{\Lambda w})-\frac{1+\beta}{2}\parallel \mathbf{\Lambda w}\parallel^2-\langle\ell,\mathbf{w}\rangle\Big\}.$$

In

$$(1+\beta)\Big\{-(\mathcal{A}\Lambda\mathbf{v},\Lambda\mathbf{w})-\frac{\mathbf{1}+\beta}{\mathbf{2}}\parallel\Lambda\mathbf{w}\parallel^2-\langle\ell,\mathbf{w}\rangle\Big\}.$$

$\mathbf{w}$ is an arbitrary function in $\mathbf{V_0}$. We may replace

$$\mathbf{w}\quad\text{by}\quad\frac{\mathbf{w}}{\mathbf{1}+\beta}.$$

Such a replacement leads to the **Minorant $\mathbf{M}_\ominus(\mathbf{v},\mathbf{w})$** that gives a lower bound of the deviation from exact solution:

For any $\mathbf{w}\in\mathbf{V_0}$,

$$\frac{\mathbf{1}}{\mathbf{2}}\parallel\Lambda(\mathbf{v}-\mathbf{u})\parallel^2\geq-\frac{\mathbf{1}}{\mathbf{2}}\parallel\Lambda\mathbf{w}\parallel^2-(\mathcal{A}\Lambda\mathbf{v},\Lambda\mathbf{w})-\langle\ell,\mathbf{w}\rangle \tag{5.11}$$

Minorant is sharp

It is easy to see that

$$\sup_{\mathbf{w}\in\mathbf{V}_0} \mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w}) = \frac{1}{2} \| \boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u}) \|^2 .$$

Indeed, take $\mathbf{w} = \mathbf{u} - \mathbf{v}$.

$$\mathbf{M}_{\ominus}(\mathbf{v}, \mathbf{u}-\mathbf{v}) = -\frac{1}{2} \| \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) \|^2 - (\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{v}, \boldsymbol{\Lambda}\mathbf{u}-\mathbf{v}) - \langle \ell, \mathbf{u}-\mathbf{v}\rangle.$$

Represent the last two terms as follows:

$$\begin{aligned}
&-(\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{v}, \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) - \langle \ell, \mathbf{u}-\mathbf{v}\rangle = \\
&= -(\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{v}, \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) + (\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}, \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) = \\
&= (\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}), \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})) = \| \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) \|^2
\end{aligned}$$

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,

$$\mathbf{M}_\ominus(\mathbf{u},\mathbf{w}) = -\frac{1}{2} \| \boldsymbol{\Lambda}\mathbf{w} \|^2 - (\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}, \boldsymbol{\Lambda}\mathbf{w}) - \langle \ell, \mathbf{w} \rangle = -\frac{1}{2} \| \boldsymbol{\Lambda}\mathbf{w} \|^2$$

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 $\widehat{\mathbf{y}} = \boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}$. Then,

$$\mathbf{M}_\oplus(\mathbf{u},\beta,\widehat{\mathbf{y}}) = (\mathbf{1}+\beta)\mathbf{D}(\boldsymbol{\Lambda}\mathbf{u},\widehat{\mathbf{y}}) + \frac{\mathbf{1}+\beta}{\mathbf{2}\beta}\,\mathbf{c}^2\|\ell + \boldsymbol{\Lambda}^*\boldsymbol{\mathcal{A}}\boldsymbol{\Lambda}\mathbf{u}\|^2 = \mathbf{0}.$$

Thus,

$$\inf_{\mathbf{y}\in\mathbf{Y}^*} \mathbf{M}_\oplus(\mathbf{u},\beta,\mathbf{y}) = \mathbf{0}.$$

Estimates of deviations in terms of the dual variable

In many cases, error estimates in terms of the dual variable (that may represent "flux" or "stress") is as important as the error control of the primal 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 $\mathbf{y} \in \mathbf{Y}^*$ be an approximation of $\mathbf{p}$. For any $\mathbf{q} \in \mathbf{Q}^*_\ell$, we obtain (from the triangle inequality and Young inequalities with $\gamma > 0$)

$$\|\, \mathbf{y} - \mathbf{p} \,\|_*^2 \leq (1+\gamma) \,\|\, \mathbf{y} - \mathbf{q} \,\|_*^2 + \left(1 + \frac{1}{\gamma}\right) \|\, \mathbf{q} - \mathbf{p} \,\|_*^2.$$

Recall that (see (5.5)) $\|\, \mathbf{q} - \mathbf{p} \,\|_*^2 \leq 2\,(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q}))$.

Therefore,

$$\| \mathbf{y} - \mathbf{p} \|_*^2 \leq (1+\gamma) \| \mathbf{y} - \mathbf{q} \|_*^2 + 2\left(1 + \frac{1}{\gamma}\right)(\mathbf{J}(\mathbf{u}) - \mathbf{I}^*(\mathbf{q})) \leq$$

$$\leq (1+\gamma) \| \mathbf{y} - \mathbf{q} \|_*^2 + 2\left(1 + \frac{1}{\gamma}\right)(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q})) =$$

Recall that

$$\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q}) \leq (1+\beta)\mathbf{D}(\Lambda\mathbf{v}, \mathbf{y}) + \left(1 + \frac{1}{\beta}\right)\frac{1}{2}\mathbf{d}_\ell{}^2(\mathbf{y})$$

so that the right–hand side is estimated by

$$(1+\gamma)\left(1 + \frac{1}{\gamma} + \frac{1}{\beta\gamma}\right)\mathbf{d}_\ell{}^2 + 2(1+\beta)\left(1 + \frac{1}{\gamma}\right)\mathbf{D}(\Lambda\mathbf{v}, \mathbf{y}).$$

Therefore,

$$\begin{aligned}\frac{1}{2} \,|\!|\!|\, \mathbf{y} - \mathbf{p} \,|\!|\!|_*^2 \quad &\leq (1+\gamma)\left(1 + \frac{1}{\gamma} + \frac{1}{\beta\gamma}\right) [\![\, \ell + \Lambda^* \mathbf{y} \,]\!]^2 + \\ &+ (1+\beta)\left(1 + \frac{1}{\gamma}\right) \mathbf{D}(\Lambda \mathbf{v}, \mathbf{y}). \qquad (5.12)\end{aligned}$$

Rewrite this estimate as follows:

$$\frac{1}{2} \,|\!|\!|\, \mathbf{y} - \mathbf{p} \,|\!|\!|_*^2 \leq \mathbf{M}_\oplus^*(\mathbf{y}, \mathbf{v}, \beta, \gamma),$$

where $\mathbf{M}_\oplus^*$ denotes the right-hand side of (5.12). This estimate holds for any $\mathbf{y} \in \mathbf{Y}^*$, positive parameters β, γ, and any $v \in \mathbf{V_0} + \mathbf{u_0}$. Here $\mathbf{v}$ is a "free" function in $\mathbf{V_0} + \mathbf{u_0}$. This "freedom" can be used to make the estimate sharper.

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 $\{\mathbf{Y}^*_\mathbf{i}\}_{\mathbf{i}=\mathbf{1}}^{\infty}$ and $\{\mathbf{V_{0i}}\}_{\mathbf{i}=\mathbf{1}}^{\infty}$ be two sequences of finite-dimensional subspaces that are dense in $\mathbf{Q}^*$ and $\mathbf{V_0}$, respectively, i.e., for any given $\varepsilon > 0$ and arbitrary elements $\mathbf{y} \in \mathbf{Y}^*$ and $\mathbf{w} \in \mathbf{V_0}$, one can find a natural number $\mathbf{k}_\varepsilon$ such that

$$\inf_{\tilde{\mathbf{w}} \in \mathbf{V_{0i}}} \|\tilde{\mathbf{w}} - \mathbf{w}\|_\mathbf{V} \leq \varepsilon, \quad \inf_{\tilde{\mathbf{y}} \in \mathbf{Y}^*_\mathbf{i}} \|\!| \tilde{\mathbf{y}} - \mathbf{y} \|\!|_{\mathbf{Q}^*} \leq \varepsilon, \quad \forall\, \mathbf{i} \geq \mathbf{k}_\varepsilon.$$

Let us show that sequences of two-sided bounds converging to the actual error can be evaluated by minimizing the Majorant on $\{\mathbf{Y}^*_\mathbf{i}\}$ 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

$$\|\mathbf{w_k} - (\mathbf{u} - \mathbf{v})\|_{\mathbf{V}} \le \varepsilon, \qquad \| \mathbf{p_k} - \mathbf{p} \|_{\mathbf{Q}^*} \le \varepsilon.$$

Define two quantities defined by solving **finite–dimensional problems**, namely

$$\mathbf{M}^{\mathbf{k}}_{\oplus} = \inf_{\substack{\mathbf{y_k} \in \mathbf{Y^*_k} \\ \boldsymbol{\beta} \in \mathbb{R}_+}} \mathbf{M}_{\oplus}(\mathbf{v}, \boldsymbol{\beta}, \mathbf{y_k}), \quad \mathbf{M}^{\mathbf{k}}_{\ominus} = \sup_{\mathbf{w_k} \in \mathbf{V_{0k}}} \mathbf{M}_{\ominus}(\mathbf{v}, \mathbf{w_k}).$$

By the definition

$$\mathbf{M}_{\ominus}(\mathbf{v}, \mathbf{w_k}) \le \mathbf{M}^{\mathbf{k}}_{\ominus} \le \frac{1}{2} \| \mathbf{u} - \mathbf{v} \|^2 \le \mathbf{M}^{\mathbf{k}}_{\oplus} \le \mathbf{M}_{\oplus}(\mathbf{v}, \boldsymbol{\beta}, \mathbf{p_k}).$$

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

$$\mathbf{M}^{\mathbf{k}}_{\oplus} \longrightarrow \frac{1}{2}\|\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\|^2,$$
$$\mathbf{M}^{\mathbf{k}}_{\ominus} \longrightarrow \frac{1}{2}\|\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\|^2$$

as the dimensionality $\mathbf{k}$ tends to $+\infty$.

Consider the **upper estimates**.

$$\mathbf{M}_{\oplus}(\mathbf{v},\beta,\mathbf{p_k}) = (1+\beta)\mathbf{D}(\boldsymbol{\Lambda}\mathbf{v},\mathbf{p_k}) + \frac{1+\beta}{2\beta}\mathbf{c}^2\|\,\ell+\boldsymbol{\Lambda}^*\mathbf{p_k}\,\|^2 =$$

$$= (1+\beta)\mathbf{D}(\boldsymbol{\Lambda}\mathbf{v},\mathbf{p_k}) + \frac{1+\beta}{2\beta}\mathbf{c}^2\|\boldsymbol{\Lambda}^*(\mathbf{p_k}-\mathbf{p})\|^2.$$

Here

$$\mathbf{D}(\boldsymbol{\Lambda}\mathbf{v},\mathbf{p_k}) = \frac{1}{2}(\boldsymbol{\Lambda}\mathbf{v}-\mathcal{A}^{-1}\mathbf{p_k},\mathcal{A}\boldsymbol{\Lambda}\mathbf{v}-\mathbf{p_k}) =$$

$$= \frac{1}{2}\Big(\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})-\mathcal{A}^{-1}(\mathbf{p_k}-\mathbf{p}),\mathcal{A}\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})-(\mathbf{p_k}-\mathbf{p})\Big) =$$

$$= \frac{1}{2}\,\|\,\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\,\|^2 + \|\,\mathbf{p_k}-\mathbf{p}\,\|_*^2 - (\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u}),\mathbf{p_k}-\mathbf{p}).$$

From the latter estimate we see that

$$\mathbf{D}(\boldsymbol{\Lambda}\mathbf{v},\mathbf{p_k}) \leq \frac{1}{2}\,\|\,\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\,\|^2 + \varepsilon\,\|\,\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\,\| + \frac{1}{2}\varepsilon^2 \quad (5.13)$$

Since

$$\|\mathbf{\Lambda}^*(\mathbf{p_k}-\mathbf{p})\|_{\mathbf{Q}^*} \;\leq\; \varepsilon,$$

we find that

$$\mathbf{M}^{\mathbf{k}}_{\oplus} \leq \mathbf{M}_{\oplus}(\mathbf{v},\varepsilon,\mathbf{p_k}) =$$

$$= (1+\varepsilon)\Big(\frac{1}{2} \parallel \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \parallel^2 + \varepsilon \parallel \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \parallel + \frac{1}{2}\varepsilon^2\Big) + \frac{1+\varepsilon}{2\varepsilon}\mathbf{c}^2\varepsilon^2 =$$

$$= \frac{1}{2} \parallel \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \parallel^2 + \mathbf{c_4}\varepsilon + \mathbf{o}(\varepsilon^2).$$

where $\mathbf{c_4} = \frac{1}{2}\left(\mathbf{c} + 2 \parallel \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \parallel + \parallel \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \parallel^2\right)$. Thus, we conclude that

$$\mathbf{M}^{\mathbf{k}}_{\oplus} \longrightarrow \frac{1}{2} \parallel \mathbf{\Lambda}(\mathbf{v}-\mathbf{u}) \parallel^2 \quad \text{as} \quad \mathbf{k} \to \infty.$$

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**.

$$\mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w_k}) = -\frac{1}{2} \parallel \boldsymbol{\Lambda}\mathbf{w_k} \parallel^2 - (\mathcal{A}\boldsymbol{\Lambda}\mathbf{v},\boldsymbol{\Lambda}\mathbf{w_k}) - \langle \ell, \mathbf{w_k}\rangle =$$
$$= -\frac{1}{2} \parallel \boldsymbol{\Lambda}\mathbf{w_k} \parallel + (\mathcal{A}\boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}),\boldsymbol{\Lambda}\mathbf{w_k}) =$$
$$= \frac{1}{2} \parallel \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) \parallel^2 - \frac{1}{2} \parallel \boldsymbol{\Lambda}(\mathbf{w_k}-(\mathbf{u}-\mathbf{v})) \parallel^2 \geq$$
$$\geq \frac{1}{2} \parallel \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) \parallel^2 - \frac{1}{2}\mathbf{c_2} \left\|\boldsymbol{\Lambda}(\mathbf{w_k}-(\mathbf{u}-\mathbf{v}))\right\|^2 .$$

This implies the estimate

$$\frac{1}{2} \parallel \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) \parallel^2 \geq \mathbf{M}^{\mathbf{k}}_{\ominus} \geq \frac{1}{2} \left\|\boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v})\right\|^2 - \mathbf{c_5}\varepsilon^2,$$

where $\mathbf{c_5} > \mathbf{0}$ depends on the norm of $\boldsymbol{\Lambda}$. Thus,

$$\mathbf{M}^{\mathbf{k}}_{\ominus} \to \frac{1}{2} \parallel \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{v}) \parallel^2 \quad \text{as } \mathbf{k} \to \infty.$$

Computable upper bound of the effectivity index

Having $\mathbf{M}_{\oplus}^{\mathbf{k}}$ and $\mathbf{M}_{\ominus}^{\mathbf{k}}$, one can define the number

$$\eta_{\mathbf{k}} := \frac{\mathbf{M}_{\oplus}^{\mathbf{k}}}{\mathbf{M}_{\ominus}^{\mathbf{k}}} \geq \mathbf{1}, \tag{5.14}$$

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

$\mathbf{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{y})$ 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 $\langle \mathbf{l}, \mathbf{w} \rangle = (\mathbf{g}, \mathbf{w})$, where $\mathbf{g} \in \mathbf{U}$, so that $\mathbf{p} \in \mathbf{Q}^* \subset \mathbf{Q}^*_{\mathbf{l}}$ and

$$\mathbf{Q}^*_\ell := \{\mathbf{y} \in \mathbf{Q}^* \mid (\mathbf{\Lambda}^* \mathbf{y} + \mathbf{g}, \mathbf{w}) = \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0}\}.$$

First, we select $\mathbf{y}$ as follows

$$\mathbf{y}^*_{\mathbf{1}} = \mathcal{A} \mathbf{\Lambda} \mathbf{v}. \tag{5.15}$$

Other variants arise if we set

$$\mathbf{y} = \mathbf{\Pi} \mathbf{y}^*_{\mathbf{1}}, \tag{5.16}$$

where $\mathbf{\Pi}$ is a certain continuous mapping.

Residual based estimate

If $\mathbf{\Pi}$ is the identity mapping of $\mathbf{Y}^*$, i.e., $\mathbf{y} = \mathbf{y_0^*}$, then

$$\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y_0^*}) = \mathbf{0}.$$

Use the majorant in the form (5.9):

$$\frac{\mathbf{1}}{\mathbf{2}} \| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^{\mathbf{2}} \leq (\mathbf{1} + \beta)\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) + \frac{\mathbf{1} + \beta}{\mathbf{2}\beta} [\![\ell + \mathbf{\Lambda}^* \mathbf{y}]\!]^{\mathbf{2}}.$$

Now, it contains only the second term, which after the minimization with respect to β gives

$$\| \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \|^{\mathbf{2}} \leq [\![\ell + \mathbf{\Lambda}^* \mathcal{A} \mathbf{\Lambda v}]\!] = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{(\mathbf{g}, \mathbf{w}) + (\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda w})}{\| \mathbf{\Lambda w} \|}. \quad (5.17)$$

If $\mathbf{v}$ is obtained by FEM and $\mathbf{v} = \mathbf{u_h} \in \mathbf{V_h} := \mathbf{V_{0h}} + \mathbf{u_0}$, (5.17) is estimated by using Galerkin orthogonality.

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 involve 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

$$\mathbf{y}_1^* = \mathbf{\Pi}_1\mathbf{y}_0^* \in \mathbf{Q}^* \tag{5.18}$$

and the quantity $\mathbf{M}_{\oplus}(\mathbf{v},\beta,\mathbf{y}_1^*)$, which leads to the error majorant

$$\mathbf{M}_{\oplus}^{(1)}(\mathbf{v}) = \inf_{\beta\in\mathbb{R}_+}\Big\{(\mathbf{1}+\beta)\mathbf{D}(\mathbf{\Lambda v},\mathbf{\Pi_1}(\mathcal{A}\mathbf{\Lambda v}))+ \\ +\frac{\mathbf{1}+\beta}{\mathbf{2}\beta}\mathbf{c}^2\,\|\,\ell+\mathbf{\Lambda}^*\mathbf{\Pi_1}(\mathcal{A}\mathbf{\Lambda v})\,\|^2\Big\}. \tag{5.19}$$

Particular case

In the simplest case associated with the problem

$$\mathbf{\Delta u + f = 0}, \qquad \mathbf{u = u_0} \quad \text{on } \partial\mathbf{\Omega}$$

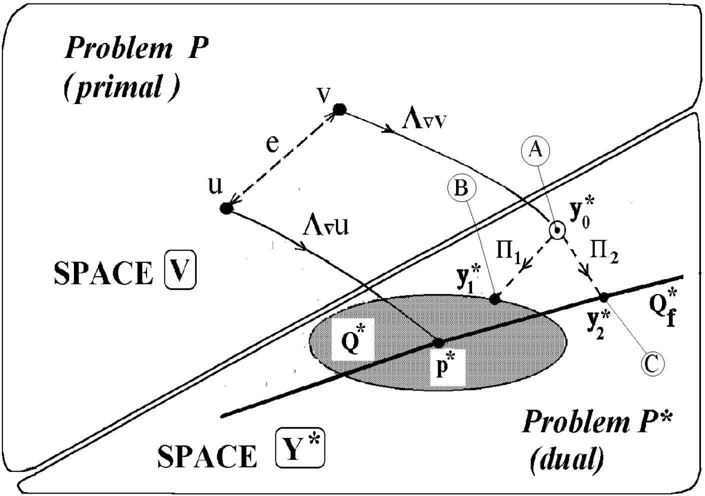
we have

$$\mathbf{M}_{\oplus}^{(1)}(\mathbf{u_h}) =$$

$$= \inf_{\beta \in \mathbb{R}_+} \left\{ (\mathbf{1}+\beta)\|\nabla \mathbf{u_h} - \mathbf{\Pi_1}(\nabla \mathbf{u_h})\|^2 + \frac{(\mathbf{1}+\beta)\mathbf{C}_{\Omega}^2}{\mathbf{2}\beta}\|\mathbf{f} + \mathbf{div}\mathbf{\Pi_1}(\nabla \mathbf{u_h})\|^2 \right\}.$$

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.

Diagram that shows connections with other methods



Estimates based on the "equilibration" of the dual variable

Let $\mathbf{\Pi_2}$ maps $\mathbf{Y}^*$ to the set $\mathbf{Q}^*_\ell$. Define

$$\mathbf{y}_2^* = \mathbf{\Pi}_2 \mathbf{y}_0^* \in \mathbf{Q}^*_\ell. \tag{5.20}$$

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 $\widetilde{\mathbf{\Pi}}_2$, which provides an approximate "equilibration". In this case, the **second term of the Majorant does not vanish and should be taken into account.**

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

$$\| \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{u_h}) \| \leq 2(1+\beta)\mathbf{D}(\boldsymbol{\Lambda}\mathbf{u_h},\mathbf{y}) + \left(1+\frac{1}{\beta}\right) [\![\boldsymbol{\Lambda}^*\mathbf{y}+\ell]\!]^2$$

Set here $\mathbf{y} = \mathcal{A}\boldsymbol{\Lambda}\mathbf{v_h}$, where $\mathbf{v_h}$ is an arbitrary element of $\mathbf{V_h}$. Then,

$$[\![\boldsymbol{\Lambda}^*\mathbf{y}+\ell]\!] = \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{(\mathbf{y}-\mathbf{p},\boldsymbol{\Lambda}\mathbf{w})}{\| \boldsymbol{\Lambda}\mathbf{w} \|} =$$

$$= \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{(\mathcal{A}\boldsymbol{\Lambda}(\mathbf{v_h}-\mathbf{u}),\boldsymbol{\Lambda}\mathbf{w})}{\| \boldsymbol{\Lambda}\mathbf{w} \|} \leq \| \boldsymbol{\Lambda}(\mathbf{v_h}-\mathbf{u}) \| .$$

It is easy to see that

$$\mathbf{D}(\mathbf{\Lambda u_h}, \mathcal{A}\mathbf{\Lambda v_h}) = \mathbf{J}(\mathbf{v_h}) - \mathbf{J}(\mathbf{u_h}).$$

Indeed,

$$\begin{aligned}\mathbf{D}(\mathbf{\Lambda u_h}, \mathcal{A}\mathbf{\Lambda v_h}) = \frac{1}{2}(\mathcal{A}\mathbf{\Lambda v_h}, \mathbf{\Lambda v_h}) + \langle \ell, \mathbf{v_h}\rangle - \\ - \frac{1}{2}(\mathcal{A}\mathbf{\Lambda u_h}, \mathbf{\Lambda u_h}) - \langle \ell, \mathbf{u_h}\rangle + \\ + (\mathcal{A}\mathbf{\Lambda u_h}, \mathbf{\Lambda}(\mathbf{u_h} - \mathbf{v_h})) + \langle \ell, \mathbf{u_h} - \mathbf{v_h}\rangle.\end{aligned}$$

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

$$\begin{aligned}\| \mathbf{\Lambda}(\mathbf{u_h} - \mathbf{u}) \|^2 = 2(\mathbf{J}(\mathbf{u_h}) - \mathbf{J}(\mathbf{u})),\\ \| \mathbf{\Lambda}(\mathbf{v_h} - \mathbf{u}) \|^2 = 2(\mathbf{J}(\mathbf{v_h}) - \mathbf{J}(\mathbf{u})).\end{aligned}$$

Therefore,

$$\mathbf{2D}(\boldsymbol{\Lambda}\mathbf{u_h}, \mathcal{A}\boldsymbol{\Lambda}\mathbf{v_h}) = \mathbf{2}(\mathbf{J}(\mathbf{v_h}) - \mathbf{J}(\mathbf{u})) - \mathbf{2}(\mathbf{J}(\mathbf{u_h}) - \mathbf{J}(\mathbf{u})) = \\ = \| \boldsymbol{\Lambda}(\mathbf{v_h} - \mathbf{u}) \|^2 - \| \boldsymbol{\Lambda}(\mathbf{u_h} - \mathbf{u}) \|^2 .$$

Now, the error estimate comes in the form

$$\| \boldsymbol{\Lambda}(\mathbf{u}-\mathbf{u_h}) \| \leq (\mathbf{1}+\beta)(\| \boldsymbol{\Lambda}(\mathbf{v_h} - \mathbf{u}) \|^2 - \| \boldsymbol{\Lambda}(\mathbf{u_h} - \mathbf{u}) \|^2) + \\ + \left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right) \| \boldsymbol{\Lambda}(\mathbf{v_h} - \mathbf{u}) \|^2 .$$

Thus, we obtain

$$(\mathbf{2} + \beta) \| \boldsymbol{\Lambda}(\mathbf{u} - \mathbf{u_h}) \|^2 \leq \\ \leq (\mathbf{1} + \beta) \| \boldsymbol{\Lambda}(\mathbf{v_h} - \mathbf{u}) \|^2 + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \| \boldsymbol{\Lambda}(\mathbf{v_h} - \mathbf{u}) \|^2,$$

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

$$\|\!| \mathbf{\Lambda}(\mathbf{u}-\mathbf{u_h}) \|\!| \leq \inf_{\mathbf{v_h}\in\mathbf{V_h}} \|\!| \mathbf{\Lambda}(\mathbf{u}-\mathbf{v_h}) \|\!| .$$

Finally, we note that functional a posteriori estimates also imply a projection type error estimate of a different type.
Let us set $\mathbf{v} = \mathbf{u_h}$, $\mathbf{y} = \mathbf{y_h} := \mathcal{A}\nabla \mathbf{u_h}$. Since

$$\mathbf{D}(\boldsymbol{\Lambda}\mathbf{u_h}, \mathbf{y_h}) = \mathbf{0},$$

we have

$$|\!|\!| \boldsymbol{\Lambda}(\mathbf{u_h} - \mathbf{u}) |\!|\!|^2 \leq |\!|\!| \mathbf{y_h} - \mathbf{q} |\!|\!|_*^2 \qquad \forall \mathbf{q} \in \mathbf{Q}^*_\ell .$$

From here, it follows the estimate

$$\| \boldsymbol{\Lambda}(\mathbf{u} - \mathbf{u_h}) \| \leq \inf_{\mathbf{q} \in \mathbf{Q}^*_\ell} \| \mathbf{y_h} - \mathbf{q} \|_*,$$

which is in a sense **dual** to the first one. It shows that an upper bound of the error is also given by the distance in the space $\mathbf{Y}^*$ between the "Galerkin flux" $\mathcal{A}\nabla\mathbf{u_h}$ and the set $\mathbf{Q}^*_\ell$ that contains the solution of the dual problem.

Lecture 6.
FUNCTIONAL A POSTERIORI ESTIMATES. LINEAR ELLIPTIC PROBLEMS.

Main goal of the lecture

In the previous lecture we have analyzed the abstract linear problem of the form

$$\Lambda^* \mathcal{A} \Lambda \mathbf{u} + \ell = 0$$

and obtained an estimate

$$\frac{1}{2} \||\, \Lambda(\mathbf{v} - \mathbf{u})\, \||^2 \leq (1+\beta) D(\Lambda \mathbf{v}, \mathbf{y}) + \frac{1+\beta}{2\beta} [\![\, \ell + \Lambda^* \mathbf{y} \,]\!]^2 .$$

In the present lecture, we discuss particular forms of this general estimate for some elliptic type boundary–value problems.

Lecture plan

- Diffusion equation with Dirichlét boundary conditions;
- Diffusion equation with Neumann boundary conditions;
- Diffusion equation with mixed boundary conditions;
- Linear elasticity with mixed boundary conditions;

Diffusion equation

Let $\mathcal{A}$ is produced by a matrix $\mathbf{A} = \{\mathbf{a_{ij}}\} = \{\mathbf{a_{ji}}\}$, $\mathbf{V} = \mathbf{H^1}(\boldsymbol{\Omega})$, where Ω is a Lipschitz domain, $\mathbf{U} = \mathbf{L^2}(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}})$, and $\boldsymbol{\Lambda}\mathbf{w} = \nabla \mathbf{w}$. Let the entries of $\mathbf{A}$ be bounded at almost all points of Ω and such that

$$\mathbf{c_1}|\xi|^2 \leq \mathbf{a_{ij}}\xi_{\mathbf{i}}\xi_{\mathbf{j}} \leq \mathbf{c_2}|\xi|^2, \qquad \forall \xi \in \mathbb{R}^{\mathbf{n}}. \tag{6.1}$$

Then, the spaces $\mathbf{Y}$ and $\mathbf{Y}^*$ have the norms

$$\|\!| \mathbf{y} \|\!|^2 = \int_{\boldsymbol{\Omega}} \mathbf{A}\mathbf{y} \cdot \mathbf{y}\, \mathbf{dx}, \quad \|\!| \mathbf{y} \|\!|_*^2 = \int_{\boldsymbol{\Omega}} \mathbf{A}^{-1}\mathbf{y} \cdot \mathbf{y}\, \mathbf{dx}.$$

Dirichlet boundary conditions

We begin with the problem

$$\mathbf{div A \nabla u = f} \quad \text{in} \quad \mathbf{\Omega}, \tag{6.2}$$

$$\mathbf{u = u_0} \quad \text{on} \quad \partial\mathbf{\Omega}. \tag{6.3}$$

In this case, $\mathbf{V_0} = \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\mathbf{\Omega})$ and $\mathbf{u}$ meets the integral identity

$$\int_{\mathbf{\Omega}} \mathbf{A\nabla u \cdot \nabla w \, dx} + \langle \mathbf{f}, \mathbf{w} \rangle = \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0}. \tag{6.4}$$

The relation $(\mathbf{y}, \mathbf{\Lambda w}) = \langle \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle$ has the form

$$\int_{\mathbf{\Omega}} \mathbf{y \cdot \nabla w \, dx} = \langle -\mathbf{div\, y}, \mathbf{w} \rangle,$$

where $\mathbf{\Lambda}^* = -\mathbf{div}$ and $\mathbf{div\, y}$ is in $\mathbf{H}^{-\mathbf{1}}(\mathbf{\Omega})$.

The operator $\boldsymbol{\Lambda}$ satisfies the required inequality

$$\mathbf{c}_{\boldsymbol{\Omega}}\|\nabla\mathbf{w}\| \geq \|\mathbf{w}\|, \qquad \forall\mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\boldsymbol{\Omega}).$$

Upper estimates of $\|\!|\ \mathbf{v}-\mathbf{u}\ |\!\|$ for an approximation $v \in \mathbf{V_0}+\mathbf{u_0}$ follow from the general estimate presented in Lecture 5. We have

$$\frac{\mathbf{1}}{\mathbf{2}}\int_{\boldsymbol{\Omega}} \mathbf{A}\nabla(\mathbf{v}-\mathbf{u})\cdot\nabla(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \leq \mathbf{M}_{\oplus}(\mathbf{v},\boldsymbol{\beta},\mathbf{y}),$$

where

$$\mathbf{M}_{\oplus}(\mathbf{v},\boldsymbol{\beta},\mathbf{y}) =$$
$$\frac{\mathbf{1}+\boldsymbol{\beta}}{\mathbf{2}}\int_{\boldsymbol{\Omega}}\left(\nabla\mathbf{v}-\mathbf{A}^{-\mathbf{1}}\mathbf{y}\right)\cdot(\,\mathbf{A}\nabla\mathbf{v}-\mathbf{y}\,)\,\mathbf{dx}+\frac{\mathbf{1}+\boldsymbol{\beta}}{\mathbf{2}\boldsymbol{\beta}}\frac{\mathbf{c}^{\mathbf{2}}_{\boldsymbol{\Omega}}}{\mathbf{c}^{\mathbf{2}}_{\mathbf{1}}}\|\mathbf{div}\,\mathbf{y}-\mathbf{f}\,\|^{\mathbf{2}} \tag{6.5}$$

Certainly, the above estimate is aplicable for the case $\mathbf{f} \in \mathbf{L}^2(\mathbf{\Omega})$ so that

$$\langle \mathbf{f}, \mathbf{w} \rangle = \int_{\mathbf{\Omega}} \mathbf{f} \mathbf{w} \, \mathbf{dx},$$

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

Let $\{\mathbf{Y}^*_{\mathbf{k}}\}$ be finite-dimensional subspaces of $\mathbf{Y}^*$ such that

$$\begin{aligned} &\mathbf{Y}^*_{\mathbf{k}} \in \mathbf{H}(\mathbf{\Omega}, \mathbf{div}) \quad \text{for all } \mathbf{k} = \mathbf{1}, \mathbf{2}, \ldots; \\ &\dim \mathbf{Y}^*_{\mathbf{k}} \to +\infty \quad \text{as } \mathbf{k} \to \infty. \end{aligned}$$

We obtain computable upper bounds

$$\mathbf{M}^{\mathbf{k}}_{\oplus} = \inf_{\substack{\mathbf{y} \in \mathbf{Y}^*_{\mathbf{k}} \\ \boldsymbol{\beta} \in \mathbb{R}_+}} \left\{ \frac{\mathbf{1}+\boldsymbol{\beta}}{\mathbf{2}} \int_{\mathbf{\Omega}} (\nabla \mathbf{v} - \mathbf{A}^{-1} \mathbf{y}) \cdot (\mathbf{A} \nabla \mathbf{v} - \mathbf{y}) \, \mathbf{dx} + \right.$$

$$\left. + \frac{1+\beta}{2\beta} \frac{\mathbf{c}^2_{\mathbf{\Omega}}}{\mathbf{c_1}} \| \, \mathbf{div}\, \mathbf{y} - \mathbf{f} \, \|^2_{\mathbf{\Omega}} \right\}. \quad (6.6)$$

Lower estimates follow

We have

$$\frac{1}{2}\int_\Omega \mathbf{A}\nabla(\mathbf{v}-\mathbf{u})\cdot\nabla(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \geq \mathbf{M}_\ominus(\mathbf{v},\mathbf{w}), \quad \forall \mathbf{w}\in \mathbf{V_0},$$

where

$$\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}) = -\frac{1}{2}\int_\Omega \mathbf{A}\nabla\mathbf{w}\cdot\nabla\mathbf{w}\,\mathbf{dx} - \int_\Omega \mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{w}\mathbf{dx} - \langle \mathbf{f},\mathbf{w}\rangle.$$

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

$$\mathbf{M}^{\mathbf{k}}_\ominus = \sup_{\mathbf{w_k}\in\mathbf{V_{0k}}} \mathbf{M}_\ominus(\mathbf{v},\mathbf{w_k}). \tag{6.7}$$

Both sequences $\mathbf{M}_{\ominus}^{\mathbf{k}}$ and $\mathbf{M}_{\oplus}^{\mathbf{k}}$ tend to $\frac{1}{2}\|\!|\,\mathbf{v}-\mathbf{u}\,|\!\|^{\mathbf{2}}$ as $\mathbf{k}\to\infty$, provided that $\{\mathbf{Y}_{\mathbf{k}}^{*}\}$ and $\{\mathbf{V}_{\mathbf{0k}}\}$ possess necessary approximation properties (limit density).
Note that if $\mathbf{v}$ is a Galerkin approximation computed on $\mathbf{V}_{\mathbf{0k}}$, then $\mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w}_{\mathbf{k}})=\mathbf{0}$. This means that to obtain a sensible lower estimate in this case, one must always use a finite-dimensional subspace that is *larger* than $\mathbf{V}_{\mathbf{0k}}$.

Neumann boundary condition

Consider the Neumann boundary condition

$$\boldsymbol{\nu} \cdot \mathbf{A}\nabla\mathbf{u} + \mathbf{F} = \mathbf{0} \quad \text{on} \quad \partial\boldsymbol{\Omega}, \tag{6.8}$$

where $\boldsymbol{\nu}$ is the vector of unit outward normal to $\partial\Omega$. To apply the general scheme we set

$$\mathbf{V_0} := \left\{ \mathbf{v} \in \mathbf{H}^1(\boldsymbol{\Omega}) \;\middle|\; \int_{\boldsymbol{\Omega}} \mathbf{v}\,\mathbf{dx} = \mathbf{0} \right\}$$

and define $\boldsymbol{\Lambda}^*\mathbf{y} \in \mathbf{V}_\mathbf{0}^*$ by the relation

$$\langle \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w} \rangle = \int_{\boldsymbol{\Omega}} \mathbf{y} \cdot \nabla\mathbf{w}\,\mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

If $\mathbf{y}$ is sufficiently regular then

$$\langle \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w} \rangle = \int_{\boldsymbol{\Omega}} (-\mathbf{div}\mathbf{y})\mathbf{w}\,\mathbf{dx} + \int_{\partial\boldsymbol{\Omega}} (\mathbf{y} \cdot \boldsymbol{\nu})\mathbf{w}\mathbf{dx}.$$

Therefore, in such a case

$$\mathbf{\Lambda}^*\mathbf{y} = [-\mathbf{div}\,\mathbf{y}\,|_{\mathbf{\Omega}}; \qquad \mathbf{y}\cdot\boldsymbol{\nu}\,|_{\partial\mathbf{\Omega}}]$$

Also, we assume that **F** and **f** satisfy the equilibrium condition

$$\int_{\mathbf{\Omega}} \mathbf{f}\,\mathbf{dx} + \int\limits_{\partial\mathbf{\Omega}} \mathbf{F}\,\mathbf{dx} = \mathbf{0}.$$

Assume that $\mathbf{f} \in \mathbf{L^2}(\mathbf{\Omega})$ and $\mathbf{F} \in \mathbf{L^2}(\partial\mathbf{\Omega})$. Then the Neumann problem has a solution defined by the integral identity

$$\int_{\mathbf{\Omega}} \mathbf{A}\nabla\mathbf{u}\cdot\nabla\mathbf{w}\,\mathbf{dx} + \langle\boldsymbol{\ell},\mathbf{w}\rangle = \mathbf{0}, \quad \forall\mathbf{w}\in\mathbf{V_0},$$

where

$$\langle\boldsymbol{\ell},\mathbf{w}\rangle = \int_{\mathbf{\Omega}} \mathbf{fw}\,\mathbf{dx} + \int\limits_{\partial\mathbf{\Omega}} \mathbf{Fw}\,\mathbf{ds}.$$

In general, $[\![\, \ell + \boldsymbol{\Lambda}^* \mathbf{y} \,]\!]$ 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

$$\mathbf{y} \in \mathbf{Q}^*(\boldsymbol{\Omega}) := \{ \mathbf{y} \in \mathbf{Y}^* \mid \operatorname{\mathbf{div}} \mathbf{y} \in \mathbf{L}^2(\boldsymbol{\Omega}), \mathbf{y} \cdot \boldsymbol{\nu} \in \mathbf{L}^2(\partial \boldsymbol{\Omega}) \},$$

then

$$\langle \ell + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_{\boldsymbol{\Omega}} (\mathbf{f} - \operatorname{\mathbf{div}} \mathbf{y}) \mathbf{w} \, \mathbf{dx} + \int\limits_{\partial \boldsymbol{\Omega}} (\mathbf{F} + \mathbf{y} \cdot \boldsymbol{\nu}) \mathbf{w} \, \mathbf{ds}$$

and, therefore,

$$|\langle \ell + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle| \leq \\ \leq \| \operatorname{\mathbf{div}} \mathbf{y} - \mathbf{f} \|_{2,\boldsymbol{\Omega}} \| \mathbf{w} \|_{2,\boldsymbol{\Omega}} + \| \mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F} \|_{2,\partial\boldsymbol{\Omega}} \| \mathbf{w} \|_{2,\partial\boldsymbol{\Omega}}. \quad (6.9)$$

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, (6.9) implies the estimate

$$|\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle| \leq$$

$$\leq \mathbf{c}_{(\Omega,\partial\Omega)} \left(\|\,\mathbf{div}\,\mathbf{y} - \mathbf{f}\,\|^2_{2,\Omega} + \|\mathbf{y}\cdot\boldsymbol{\nu} + \mathbf{F}\|^2_{2,\partial\Omega} \right)^{1/2} \interleave\, \boldsymbol{\Lambda}\mathbf{w}\, \interleave^2$$

and the second term of the majorant is calculated as follows:

$$[\![\, \boldsymbol{\ell} + \boldsymbol{\Lambda}^*\mathbf{y} \,]\!] = \sup_{\mathbf{w}\in\mathbf{V}_0} \frac{\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle}{\interleave\, \boldsymbol{\Lambda}\mathbf{w}\, \interleave} \leq$$

$$\leq \mathbf{c}_{(\Omega,\partial\Omega)} \left(\|\,\mathbf{div}\,\mathbf{y} - \mathbf{f}\,\|^2_{2,\Omega} + \|\mathbf{y}\cdot\boldsymbol{\nu} + \mathbf{F}\|^2_{2,\partial\Omega} \right)^{1/2}.$$

The term $\mathbf{D}(\boldsymbol{\Lambda}v, \mathbf{y})$ 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 $\mathbf{Y}^*_\mathbf{k}$, then one can compute $\mathbf{M}^\mathbf{k}_\oplus$ as

$$\mathbf{M}^\mathbf{k}_\oplus = \inf_{\substack{\mathbf{y}\in\mathbf{Y}^*_\mathbf{k},\ \mathbf{y}\cdot\nu=\mathbf{F}\ \mathrm{on}\ \partial\Omega\\ \boldsymbol{\beta}\in\mathbb{R}_+}} \left\{\frac{\mathbf{1}+\boldsymbol{\beta}}{\mathbf{2}}\int_\Omega (\nabla\mathbf{v}-\mathbf{A}^{-1}\mathbf{y})\cdot(\mathbf{A}\nabla\mathbf{v}-\mathbf{y})\,\mathbf{dx}+ \right.$$

$$\left. +\frac{1+\beta}{2\beta}\,\mathbf{c}^2_{(\Omega,\partial\Omega)}\|\,\mathbf{div}\,\mathbf{y}-\mathbf{f}\,\|^2_\Omega\right\}. \quad (6.10)$$

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:

$$\mathbf{u} = \mathbf{u_0} \quad \text{on} \quad \partial_1 \boldsymbol{\Omega},$$

$$\boldsymbol{\nu} \cdot \mathbf{A}\nabla\mathbf{u} + \mathbf{F} = \mathbf{0} \quad \text{on} \quad \partial_2 \boldsymbol{\Omega}.$$

Set

$$\mathbf{V_0} := \left\{ \mathbf{v} \in \mathbf{H}^1(\boldsymbol{\Omega}) \,|\, \mathbf{v} = \mathbf{0} \quad \text{on} \quad \partial_1 \boldsymbol{\Omega} \right\}$$

and

$$\langle \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle = \int_{\boldsymbol{\Omega}} \mathbf{y} \cdot \nabla \mathbf{w} \, \mathbf{dx}, \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

Assume that

$$\mathbf{f} \in \mathbf{L}^2(\mathbf{\Omega}), \qquad \mathbf{F} \in \mathbf{L}^2(\partial_2\mathbf{\Omega}).$$

and $\mathbf{y}$ possesses an extra regularity, namely,

$$\mathbf{y} \in \mathbf{Q}^*(\mathbf{\Omega}) := \left\{ \mathbf{y} \in \mathbf{Y}^* \,|\, \mathbf{div}\,\mathbf{y} \in \mathbf{L}^2(\mathbf{\Omega}),\ \mathbf{y}\cdot\boldsymbol{\nu} \in \mathbf{L}^2(\partial_2\mathbf{\Omega}) \right\}.$$

Then, for any $\mathbf{w} \subset \mathbf{V_0}$, we have

$$\langle \ell + \mathbf{\Lambda}^*\mathbf{y}, \mathbf{w} \rangle = \int_{\mathbf{\Omega}} (\mathbf{div}\,\mathbf{y} - \mathbf{f})\mathbf{w}\,\mathbf{dx} + \int_{\partial_2\mathbf{\Omega}} (\mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F})\mathbf{w}\,\mathbf{ds},$$

Note that $\mathbf{p} \in \mathbf{Q}^*(\mathbf{\Omega})$!

Now, we obtain

$$|\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle| \leq \|\mathbf{div y} - \mathbf{f}\|_{2,\Omega} \|\mathbf{w}\|_{2,\Omega} + \\ + \|\mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F}\|_{2,\partial_2\Omega} \|\mathbf{w}\|_{2,\partial_2\Omega}.$$

Let γ and γ_* be two numbers such that $\gamma > \mathbf{1}, \quad \gamma_* > \mathbf{1}$, $\frac{\mathbf{1}}{\gamma} + \frac{\mathbf{1}}{\gamma_*} = \mathbf{1}$. Use the algebraic inequality

$$\mathbf{ab} + \mathbf{cd} \leq \sqrt{\gamma \mathbf{a}^2 + \gamma_* \mathbf{c}^2} \sqrt{\frac{\mathbf{1}}{\gamma} \mathbf{b}^2 + \frac{\mathbf{1}}{\gamma_*} \mathbf{d}^2}.$$

Then

$$|\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle| \leq \left(\gamma \|\mathbf{div\, y} - \mathbf{f}\|^2_{2,\Omega} + \gamma_* \|\mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F}\|^2_{2,\partial_2\Omega} \right)^{\mathbf{1/2}} \times \\ \times \left(\frac{1}{\gamma} \|\mathbf{w}\|^2_{2,\Omega} + \frac{\mathbf{1}}{\gamma_*} \|\mathbf{w}\|^2_{2,\partial_2\Omega} \right)^{1/2}.$$

Since (Friederichs type inequality)

$$\|\mathbf{w}\|^2_{2,\Omega} \leq \mathbf{C}^2_{\mathbf{F}}(\Omega)\|\nabla \mathbf{w}\|^2_{2,\Omega}, \quad \forall \mathbf{w} \in \mathbf{V_0},$$

and (trace inequality)

$$\|\mathbf{w}\|^2_{2,\partial_2\Omega} \leq \mathbf{C}^2_{\mathbf{tr}}(\Omega, \partial_2\Omega)\|\mathbf{w}\|^2_{1,2,\Omega}, \quad \forall \mathbf{w} \in \mathbf{V_0},$$

we find that

$$\frac{1}{\gamma}\|\mathbf{w}\|^2_{2,\Omega} + \frac{1}{\gamma_*}\|\mathbf{w}\|^2_{2,\partial_2\Omega} \leq$$

$$\leq \mathbf{C}^2_{\mathbf{F}}\frac{1}{\gamma}\|\nabla \mathbf{w}\|^2 + \mathbf{C}^2_{\mathbf{tr}}\frac{1}{\gamma_*}\left(\|\mathbf{w}\|^2_{2,\Omega} + \|\nabla \mathbf{w}\|^2_{2,\Omega}\right) \leq$$

$$\leq \left(\mathbf{C}^2_{\mathbf{F}}\frac{1}{\gamma} + \mathbf{C}^2_{\mathbf{tr}}\frac{1}{\gamma_*}\left(1 + \mathbf{C}^2_{\mathbf{F}}\right)\right)\|\nabla \mathbf{w}\|^2_{2,\Omega}.$$

Therefore, there exist a positive constant $\mathbf{C}_\gamma$ such that

$$\frac{1}{\mathbf{C}_\gamma^2} = \inf_{\mathbf{w}\in\mathbf{V}_0} \frac{\int_\Omega \mathbf{A}\nabla\mathbf{w}\cdot\nabla\mathbf{w}\,\mathbf{dx}}{\frac{1}{\gamma}\|\mathbf{w}\|^2_{2,\Omega} + \frac{1}{\gamma_*}\|\mathbf{w}\|^2_{2,\partial_2\Omega}}.$$

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

$$\mathbf{C}_\gamma^2 \le \widehat{\mathbf{C}}_\gamma^2 := \left(\mathbf{C}_\mathbf{F}^2\frac{1}{\gamma} + \mathbf{C}_{\mathbf{tr}}^2(1+\mathbf{C}_\mathbf{F}^2)\frac{1}{\gamma_*}\right)\mathbf{c}_1^{-1},$$

so that an upper bound of$\mathbf{C}_\gamma$ is directly computed. Now,

$$|\langle \boldsymbol{\ell} + \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle| \le$$

$$\le \widehat{\mathbf{C}}_\gamma\left(\gamma\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|^2_{2,\Omega} + \gamma_*\|\mathbf{y}\cdot\boldsymbol{\nu} + \mathbf{F}\|^2_{2,\partial_2\Omega}\right)^{1/2} ||| \nabla\mathbf{w} ||| .$$

From this estimate, we obtain

$$[\![\, \ell + \Lambda^* \mathbf{y} \,]\!]^2 \le \widehat{\mathbf{C}}_\gamma^2 \Big(\gamma \| \mathbf{div}\, \mathbf{y} - \mathbf{f} \|_{2,\Omega}^2 + \frac{\gamma}{\gamma - 1} \| \mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F} \|_{2,\partial_2\Omega}^2 \Big).$$

Consider first the case, in which we simply set $\gamma = \gamma^* = 2$. Then

$$\widehat{\mathbf{C}}_{(\gamma=2)}^2 := \widehat{\mathbf{C}}_2^2 = \frac{1}{2} \left(\mathbf{C}_{\mathbf{F}}^2 + \mathbf{C}_{\mathbf{tr}}^2 (1 + \mathbf{C}_{\mathbf{F}}^2) \right) \mathbf{c}_1^{-1},$$

$$[\![\, \ell + \Lambda^* \mathbf{y} \,]\!]^2 \le 2 \widehat{\mathbf{C}}_2^2 \Big(\| \mathbf{div}\, \mathbf{y} - \mathbf{f} \|_{2,\Omega}^2 + \| \mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F} \|_{2,\partial_2\Omega}^2 \Big).$$

and we find that

$$\mathbf{M}_{\oplus}(\mathbf{v}, \beta, \mathbf{y}) = \frac{1+\beta}{2} \int_\Omega (\nabla \mathbf{v} - \mathbf{A}^{-1} \mathbf{y}) \cdot (\mathbf{A} \nabla \mathbf{v} - \mathbf{y})\, \mathbf{dx} + \\ + \frac{1+\beta}{2\beta} \widehat{\mathbf{C}}_2^2 \Big(\| \mathbf{div}\, \mathbf{y} - \mathbf{f} \|_{2,\Omega}^2 + \| \mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F} \|_{2,\partial_2\Omega}^2 \Big). \quad (6.11)$$

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 $\beta > 0$.

A more exact estimate is obtained if we define γ by minimizing of the quantity

$$\left(\mathbf{C}_{\mathbf{F}}^{2}\frac{\mathbf{1}}{\gamma}+\mathbf{C}_{\mathbf{tr}}^{2}(\mathbf{1}+\mathbf{C}_{\mathbf{F}}^{2})\frac{\mathbf{1}}{\gamma_{*}}\right)\left(\gamma\|\mathbf{div}\,\mathbf{y}-\mathbf{f}\|_{2,\Omega}^{2}+\gamma^{*}\|\mathbf{y}\cdot\boldsymbol{\nu}+\mathbf{F}\|_{2,\partial_{2}\Omega}^{2}\right)=$$

$$\mathbf{C}_{\mathbf{F}}^{2}\frac{\gamma^{*}}{\gamma}\|\mathbf{y}\cdot\boldsymbol{\nu}+\mathbf{F}\|_{2,\partial_{2}\Omega}^{2}+\frac{\gamma^{*}}{\gamma}\mathbf{C}_{\mathbf{tr}}^{2}(\mathbf{1}+\mathbf{C}_{\mathbf{F}}^{2})\|\mathbf{div}\,\mathbf{y}-\mathbf{f}\|_{2,\Omega}^{2}+\mathrm{const}(\gamma).$$

Denote

$$\boldsymbol{\rho}_1=\|\mathbf{div}\,\mathbf{y}-\mathbf{f}\|_{2,\Omega}^{2}\qquad \boldsymbol{\rho}_2=\|\mathbf{y}\cdot\boldsymbol{\nu}+\mathbf{F}\|_{2,\partial_{2}\Omega}^{2},$$

$$\kappa_{\mathbf{1}}=\mathbf{C}_{\mathbf{F}}^{2},\qquad \kappa_{\mathbf{2}}=\mathbf{C}_{\mathbf{tr}}^{2}(\mathbf{1}+\mathbf{C}_{\mathbf{F}}^{2}).$$

Then the problem is

$$\min_{\gamma}\left(\kappa_1^2\boldsymbol{\rho}_2^2\frac{1}{\gamma-1}+(\gamma-1)\kappa_2^2\boldsymbol{\rho}_1^2\right).$$

Its minimum is attained at $\widehat{\gamma}=1+\kappa_1\boldsymbol{\rho}_2/\kappa_2\boldsymbol{\rho}_1$.

In other words, we observe that the lowest estimate of the term $[\![\ell + \mathbf{\Lambda}^* \mathbf{y}]\!]$ is attained if

$$\gamma = \widehat{\gamma} := 1 + \frac{\|\mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F}\|_{\mathbf{2},\partial_{\mathbf{2}}\boldsymbol{\Omega}} \mathbf{C_F}}{\|\mathbf{div}\, \mathbf{y} - \mathbf{f}\|_{\mathbf{2},\boldsymbol{\Omega}} \mathbf{C_{tr}} (\mathbf{1} + \mathbf{C_F^2})^{\mathbf{1/2}}}.$$

Let us find the respective upper bound. We need to calculate

$$\left(\kappa_{\mathbf{1}}^{\mathbf{2}} \frac{\mathbf{1}}{\widehat{\gamma}} + \kappa_{\mathbf{2}}^{\mathbf{2}} \frac{\mathbf{1}}{\widehat{\gamma}_*}\right) (\widehat{\gamma} \boldsymbol{\rho}_{\mathbf{1}}^{\mathbf{2}} + \widehat{\gamma}^* \boldsymbol{\rho}_{\mathbf{2}}^{\mathbf{2}}) =$$

$$= \frac{\mathbf{1}}{\widehat{\gamma}} (\kappa_{\mathbf{1}}^{\mathbf{2}} + (\widehat{\gamma} - \mathbf{1}) \kappa_{\mathbf{2}}^{\mathbf{2}}) \frac{\widehat{\gamma}}{\widehat{\gamma} - \mathbf{1}} ((\widehat{\gamma} - \mathbf{1}) \boldsymbol{\rho}_{\mathbf{1}}^{\mathbf{2}} + \boldsymbol{\rho}_{\mathbf{2}}^{\mathbf{2}}) =$$

$$= \frac{\kappa_2 \boldsymbol{\rho}_1}{\kappa_1 \boldsymbol{\rho}_2} (\kappa_1^2 + \frac{\kappa_1 \boldsymbol{\rho}_2}{\kappa_2 \boldsymbol{\rho}_1} \kappa_2^2)(\boldsymbol{\rho}_1^2 \frac{\kappa_1 \boldsymbol{\rho}_2}{\kappa_2 \boldsymbol{\rho}_1} + \boldsymbol{\rho}_2^2) = \kappa_2 \boldsymbol{\rho}_1 (\kappa_1 + \frac{\boldsymbol{\rho}_2}{\boldsymbol{\rho}_1} \kappa_2)(\boldsymbol{\rho}_1 \frac{\kappa_1}{\kappa_2} + \boldsymbol{\rho}_2) =$$

$$= (\kappa_{\mathbf{1}} \boldsymbol{\rho}_{\mathbf{1}} + \boldsymbol{\rho}_{\mathbf{2}} \kappa_{\mathbf{2}})(\boldsymbol{\rho}_{\mathbf{1}} \kappa_{\mathbf{1}} + \boldsymbol{\rho}_{\mathbf{2}} \kappa_{\mathbf{2}}) = (\kappa_{\mathbf{1}} \boldsymbol{\rho}_{\mathbf{1}} + \boldsymbol{\rho}_{\mathbf{2}} \kappa_{\mathbf{2}})^{\mathbf{2}}.$$

$\mathbf{M}_\oplus$ for mixed boundary conditions

By recalling the definitions of κ_1, κ_2, $\boldsymbol{\rho}_1$, and $\boldsymbol{\rho}_2$ we obtain

$$\begin{aligned}[\![\, \ell + \Lambda^* \mathbf{y} \,]\!]^2 \quad &\leq \Big(\mathbf{C_F} \| \mathbf{div}\, \mathbf{y} - \mathbf{f} \|_{2,\Omega} + \\ &+ \mathbf{C_{tr}} (1 + \mathbf{C_F^2})^{1/2} \| \mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F} \|_{2,\partial_2 \Omega} \Big)^2 \mathbf{c}_1^{-2}\end{aligned}$$

and we have

$$\begin{aligned}\mathbf{M}_\oplus(\mathbf{v}, \beta, \mathbf{y}) = \frac{1+\beta}{2} \int_\Omega (\nabla \mathbf{v} - \mathbf{A}^{-1} \mathbf{y}) \cdot (\mathbf{A} \nabla \mathbf{v} - \mathbf{y})\, \mathbf{dx} + \\ + \frac{1+\beta}{2\beta} \Big(\mathbf{C_F} \| \mathbf{div}\, \mathbf{y} - \mathbf{f} \|_{2,\Omega} + \\ + \mathbf{C_{tr}} (1 + \mathbf{C_F^2})^{1/2} \| \mathbf{y} \cdot \boldsymbol{\nu} + \mathbf{F} \|_{2,\partial_2 \Omega} \Big)^2 \mathbf{c}_1^{-2}. \quad (6.12)\end{aligned}$$

Majorant vanishes if and only if $\mathbf{v} = \mathbf{u}$ and $\mathbf{y} = \mathbf{A} \nabla \mathbf{u}$, it is continuous with respect to the convergence of $\mathbf{v}$ in $\mathbf{V}$ and $\mathbf{y}$ in $\mathbf{Q}$.

Lower estimates

Lower estimates for the problems considered follow from the general ones obtained in the previous lecture. They have the form

$$\frac{1}{2}\int_\Omega \mathbf{A}\nabla(\mathbf{v}-\mathbf{u})\cdot\nabla(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \geq \mathbf{M}_\ominus(\mathbf{v},\mathbf{w}), \quad \forall \mathbf{w}\in\mathbf{V_0},$$

where

$$\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}) = -\frac{1}{2}\int_\Omega \mathbf{A}\nabla(\mathbf{w}-\mathbf{v})\cdot\nabla\mathbf{w}\,\mathbf{dx} - \int_\Omega \mathbf{fw}\,\mathbf{dx} - \int_{\partial_2\Omega} \mathbf{Fw}\,\mathbf{ds}.$$

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

$$\boldsymbol{\sigma}^* = \mathbb{L}\varepsilon(\mathbf{u}) \quad \text{in} \quad \boldsymbol{\Omega}, \qquad \textbf{(Hooke's law)}$$

$$\mathbf{div}\boldsymbol{\sigma}^* = \mathbf{f} \ \text{in} \quad \boldsymbol{\Omega}, \qquad \textbf{(Equilibrium equation)}$$

$$\mathbf{u} = \mathbf{u_0} \quad \text{on} \quad \partial_{\mathbf{1}}\boldsymbol{\Omega},$$

$$\boldsymbol{\sigma}^*\boldsymbol{\nu} + \mathbf{F} = \mathbf{0} \ \text{on} \quad \partial_{\mathbf{2}}\boldsymbol{\Omega}.$$

where $\varepsilon(\mathbf{u})$ 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

$$\mathbf{C_1}|\varepsilon|^{\mathbf{2}} \leq \mathbb{L}\varepsilon : \varepsilon \leq \mathbf{C_2}|\varepsilon|^{\mathbf{2}}, \quad \forall \varepsilon \in \mathbb{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}},$$

and

$$L_{ijkm} = L_{jikm} = L_{kmij}, \quad L_{ijkm} \in \mathbf{L}^{\infty}(\mathbf{\Omega}).$$

Generalized solution

Let

$$\mathbf{f} \in \mathbf{L}^2(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}}), \quad \mathbf{F} \in \mathbf{L}^2(\partial_2\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}}).$$

Then, a generalized solution $\mathbf{u} \in \mathbf{V_0} + \mathbf{u_0}$ is defined by the identity

$$\int_{\boldsymbol{\Omega}} \mathbb{L}\, \boldsymbol{\varepsilon}(\mathbf{u}) : \boldsymbol{\varepsilon}(\mathbf{w})\, \mathbf{dx} + \langle \boldsymbol{\ell}, \mathbf{w} \rangle \;=\; \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0}, \qquad (6.13)$$

where

$$\langle \boldsymbol{\ell}, \mathbf{w} \rangle = \int_{\boldsymbol{\Omega}} \mathbf{f} \cdot \mathbf{w}\, \mathbf{dx} + \int_{\partial_2\boldsymbol{\Omega}} \mathbf{F} \cdot \mathbf{w}\, \mathbf{ds}.$$

Assume that $\mathbf{u}$ is a smooth function and it satisfies the identity

$$\int_{\Omega} \mathbb{L}\,\varepsilon(\mathbf{u}) : \varepsilon(\mathbf{w})\,\mathbf{dx} + \langle \ell, \mathbf{w} \rangle \;=\; \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0},$$

Then,

$$\int_{\Omega} (\mathbf{f} - \mathbf{div}(\mathbb{L}\,\varepsilon(\mathbf{u})) \cdot \mathbf{w}\,\mathbf{dx} + \int\limits_{\partial_2\Omega} \Big((\mathbb{L}\,\varepsilon(\mathbf{u}))\boldsymbol{\nu} + \mathbf{F} \Big) \cdot \mathbf{w}\,\mathbf{ds} \;=\; \mathbf{0},$$
$$\forall \mathbf{w} \in \mathbf{V_0},$$

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 (6.13) is the Euler's equation for the functional

$$\mathbf{J}(\mathbf{v}) = \frac{1}{2}\int_{\Omega} \mathbb{L}\varepsilon(\mathbf{v}) : \varepsilon(\mathbf{v})\,\mathbf{dx} + < \ell, \mathbf{v} > .$$

Therefore, the respective boundary–value problem may be considered as a minimization problem for $\mathbf{J}(\mathbf{v})$ on the set

$$\mathbf{V_0} := \{\mathbf{v} \in \mathbf{H}^1(\Omega, \mathbb{R}^{\mathbf{n}}) \mid \mathbf{v} = \mathbf{u_0} \text{ on } \partial_1\Omega\}.$$

To prove existence of a minimizer we must show the coercivity of $\mathbf{J}(\mathbf{v})$ on $\mathbf{V_0}$. The key role in this belongs to the so–called Korn's inequality.

In the Dirichl et problem

$$\mathbf{J}(\mathbf{v}) = \frac{1}{2}\int_\Omega \mathbb{L}\varepsilon(\mathbf{v}) : \varepsilon(\mathbf{v})\,\mathbf{dx} + < \ell, \mathbf{v} > \geq$$
$$\geq \frac{\mathbf{C_1}}{2}\|\varepsilon(\mathbf{v})\|^2 - \|\mathbf{f}\|\|\mathbf{v}\| =$$
$$= \frac{\mathbf{C_1}}{2}\|\varepsilon(\mathbf{u_0}+\mathbf{w})\|^2 - \|\mathbf{f}\|\|\mathbf{u_0}+\mathbf{w}\| \geq$$
$$\geq \frac{\mathbf{C_1}}{2}(\|\varepsilon(\mathbf{u_0})\| - \|\varepsilon(\mathbf{w})\|)^2 - \|\mathbf{f}\|\|\mathbf{u_0}\| - \|\mathbf{f}\|\|\mathbf{w}\|.$$

Thus, if we can prove that

$$\|\varepsilon(\mathbf{w})\| \geq \mathbf{c}\|\nabla\mathbf{w}\| \quad \forall \mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^{\mathbf{1}}(\Omega),$$

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

$$|\!|\!|\, \mathbf{w} \,|\!|\!|_{\mathbf{1,2,\Omega}} := \left(\int_{\mathbf{\Omega}} (|\varepsilon(\mathbf{w})|^2 + |\mathbf{w}|^2)\, \mathbf{dx} \right)^{\mathbf{1/2}}.$$

Korns's inequality in $\overset{\circ}{\mathbf{H}}{}^1$

For the functions in $\overset{\circ}{\mathbf{H}}{}^1(\Omega)$ this fact is not difficult to prove. Indeed,

$$\int_\Omega |\varepsilon(\mathbf{w})|^2 d\mathbf{x} = \frac{1}{2}\|\nabla \mathbf{w}\|^2 + \frac{1}{2}\int_\Omega \sum_{ij} w_{i,j} w_{j,i} d\mathbf{x} =$$

$$= \frac{1}{2}\|\nabla \mathbf{w}\|^2 - \frac{1}{2}\int_\Omega \sum_{ij} w_i w_{j,ij} d\mathbf{x} = \frac{1}{2}\|\nabla \mathbf{w}\|^2 + \frac{1}{2}\int_\Omega \sum_{ij} w_{i,i} w_{j,j} d\mathbf{x} =$$

$$= \frac{1}{2}\|\nabla \mathbf{w}\|^2 + \frac{1}{2}\int_\Omega \sum_{i} |w_{i,i}|^2 d\mathbf{x}.$$

Thus,

$$\|\nabla \mathbf{w}\| \leq \sqrt{2}\|\varepsilon(\mathbf{w})\| \qquad \forall \mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^1(\Omega). \tag{6.14}$$

By (6.14) we prove that the energy functional of the elasticity problem for the case of Dirichl'et boundary conditions is coercive, i.e.,

$$\mathbf{J}(\mathbf{v_k}) \longrightarrow +\infty, \quad \textbf{as } \|\nabla \mathbf{v_k}\| \longrightarrow +\infty.$$

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 $\mathbf{RD}(\boldsymbol{\Omega})$. 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:

$$\mathbf{RD}(\boldsymbol{\Omega}) := \{\mathbf{w} = \mathbf{w_0} + \omega_\mathbf{0}\mathbf{x} \mid \mathbf{w_0} \in \mathbb{R}^{\mathbf{n}},\ \omega_\mathbf{0} \in \mathbb{M}^{\mathbf{n}\times\mathbf{n}}\},$$

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,

$$\|\mathbf{u}\|_{1,\mathbf{p},\Omega} \leq \mathbf{C}\left(\int\limits_{\Omega} |\varepsilon(\mathbf{u})|^{\mathbf{p}}\,\mathbf{dx}\right)^{\frac{1}{\mathbf{p}}} \quad \forall \mathbf{u} \in \mathbf{V}, \quad \mathbf{p} \in (\mathbf{1},\mathbf{2}] \quad (6.15)$$

Proof. Assume the opposite. Then, for any $m \in \mathbb{N}$ we can find $\mathbf{v}^{(\mathbf{m})}$ such that$\mathbf{v}^{(\mathbf{m})} \in \mathbf{V}$ and

$$\|\mathbf{v}^{(\mathbf{m})}\|_{1,\mathbf{p},\Omega} > \mathbf{m}\left(\int\limits_{\Omega} |\varepsilon(\mathbf{v}^{(\mathbf{m})})|^{\mathbf{p}}\,\mathbf{dx}\right)^{\frac{1}{\mathbf{p}}}.$$

Set $\mathbf{w}^{(\mathbf{m})} = \frac{\mathbf{v}^{(\mathbf{m})}}{\|\mathbf{v}^{(\mathbf{m})}\|_{\mathbf{1},\mathbf{p},\boldsymbol{\Omega}}}$, then

$$\|\mathbf{w}^{(\mathbf{m})}\|_{\mathbf{1},\mathbf{p},\boldsymbol{\Omega}} = \mathbf{1} \qquad \text{and} \qquad \frac{\mathbf{1}}{\mathbf{m}} \geq \left(\int\limits_{\boldsymbol{\Omega}} |\varepsilon(\mathbf{w}^{(\mathbf{m})}|^{\mathbf{p}} \, \mathbf{dx} \right)^{\frac{\mathbf{1}}{\mathbf{p}}} .$$

Therefore,

$$\begin{aligned} \mathbf{w}^{(\mathbf{m})} &\rightharpoonup \mathbf{w} \quad \text{in} \quad \mathbf{W}^{\mathbf{1}}_{\mathbf{p}}(\boldsymbol{\Omega}, \mathbf{R}^{\mathbf{n}}), \\ \mathbf{w}^{(\mathbf{m})} &\to \mathbf{w} \quad \in \quad \mathbf{L}^{\mathbf{p}}(\boldsymbol{\Omega}, \mathbf{R}^{\mathbf{n}}), \\ \|\varepsilon(\mathbf{w}^{(\mathbf{m})})\|_{\mathbf{p},\boldsymbol{\Omega}} &\to \mathbf{0} \quad \text{in} \quad \mathbf{L}^{\mathbf{p}}(\boldsymbol{\Omega}, \mathbf{R}^{\mathbf{n}}). \end{aligned}$$

From here we conclude that $\varepsilon(\mathbf{w}) = \mathbf{0}$.

Indeed, by the fact that a norm is weakly lower semicontinuous, we have

$$\mathbf{0} = \liminf_{\mathbf{m}} \|\varepsilon(\mathbf{w}^{(\mathbf{m})})\|_{\mathbf{p},\mathbf{\Omega}} \geq \|\varepsilon(\mathbf{w})\|_{\mathbf{p},\mathbf{\Omega}}.$$

Thus, $\mathbf{w} \in \mathbf{RD}(\mathbf{\Omega}) \cap \mathbf{V}$. 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

$$\|\mathbf{w}(\mathbf{m})\|_{\mathbf{1},\mathbf{p},\mathbf{\Omega}} \leq \mathbf{C} \left(\int\limits_{\mathbf{\Omega}} \left(|\varepsilon(\mathbf{w}^{(\mathbf{m})})|^{\mathbf{p}} + |\mathbf{w}^{(\mathbf{m})}|^{\mathbf{p}} \right) \mathbf{dx} \right)^{\frac{\mathbf{1}}{\mathbf{p}}} \underset{\mathbf{m}\to\infty}{\longrightarrow} \mathbf{0}.$$

which shows that $\|\mathbf{w}(\mathbf{m})\|_{\mathbf{1},\mathbf{p},\mathbf{\Omega}}$ tends to zero. But for any m $\|w^{(m)}\|_{p,1,\Omega} = 1$, 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

$$\mathbf{V} = \left\{ \mathbf{v} \in \mathbf{W}^1_{\mathbf{p}}(\boldsymbol{\Omega}) \mid \int\limits_{\boldsymbol{\Omega}} \mathbf{v} \cdot \mathbf{w} \, \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{RD}(\boldsymbol{\Omega}) \right\}.$$

Theorem

Let Ω be a bounded domain with Lipschitz boundary $\partial\Omega$. Then

$$\|\mathbf{u}\|_{\mathbf{1},\mathbf{p},\boldsymbol{\Omega}} \leq \mathbf{C} \left(\int\limits_{\boldsymbol{\Omega}} |\varepsilon(\mathbf{u})|^{\mathbf{p}} \, \mathbf{dx} \right)^{\frac{1}{\mathbf{p}}} \qquad \forall \mathbf{u} \in \mathbf{V}. \tag{6.16}$$

Proof. By the same arguments as before, we obtain a sequence $\mathbf{w}^{(\mathbf{m})} \in \mathbf{V}$ such that

$$\begin{aligned} \mathbf{w}^{(\mathbf{m})} &\rightharpoonup \mathbf{w} \quad \text{in} \quad \mathbf{W}^1_{\mathbf{p}}(\boldsymbol{\Omega}, \mathbf{R}^{\mathbf{n}}), \\ \mathbf{w}^{(\mathbf{m})} &\to \mathbf{w} \quad \in \quad \mathbf{L}^{\mathbf{p}}(\boldsymbol{\Omega}, \mathbf{R}^{\mathbf{n}}), \\ \|\varepsilon(\mathbf{w}^{(\mathbf{m})})\|_{\mathbf{p},\boldsymbol{\Omega}} &\to \mathbf{0} \quad \text{in} \quad \mathbf{L}^{\mathbf{p}}(\boldsymbol{\Omega}, \mathbf{R}^{\mathbf{n}}). \end{aligned}$$

By the arguments similar to those in the previous Theorem, we find that $\varepsilon(\mathbf{w}) = \mathbf{0}$ and, thus, $\mathbf{w} \in \mathbf{RD}(\boldsymbol{\Omega})$. In addition, for any $\bar{\mathbf{w}} \in \mathbf{RD}$, we have

$$\mathbf{0} = \int_{\boldsymbol{\Omega}} \mathbf{w}^{(\mathbf{m})} \cdot \bar{\mathbf{w}} \, \mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{w} \cdot \bar{\mathbf{w}} \, \mathbf{dx}.$$

But $\mathbf{w} \in \mathbf{RD}$, so that $\|\mathbf{w}\| = \mathbf{0}$, and by applying Korn's inequality we prove that $\|\mathbf{w}^{(\mathbf{m})}\|_{\mathbf{1},\mathbf{p},\boldsymbol{\Omega}}$ 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

$$\mathbf{U}=\mathbf{L}^2(\boldsymbol{\Omega},\mathbb{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}}),\qquad \mathbf{V}=\mathbf{H}^1(\boldsymbol{\Omega},\mathbb{R}^{\mathbf{n}}),$$
$$\mathbf{V_0}=\{\mathbf{w}\in\mathbf{V}\,|\,\mathbf{w}=\mathbf{0}\ \text{on}\ \partial_1\boldsymbol{\Omega}\},$$
$$\|\!|\ \mathbf{y}\ |\!\|^2=\int_{\boldsymbol{\Omega}}\mathbb{L}\mathbf{y}:\mathbf{y}\,\mathbf{dx},\qquad \|\!|\ \mathbf{y}\ |\!\|_*^2=\int_{\boldsymbol{\Omega}}\mathbb{L}^{-1}\mathbf{y}:\mathbf{y}\,\mathbf{dx},$$

and $\boldsymbol{\Lambda}\mathbf{v}=\varepsilon(\mathbf{v}):=\frac{1}{2}\left(\nabla\mathbf{v}+(\nabla\mathbf{v})^{\mathbf{T}}\right)$. In this case,

$$\langle\boldsymbol{\Lambda}^*\mathbf{y},\mathbf{w}\rangle=\int_{\boldsymbol{\Omega}}\mathbf{y}:\varepsilon(\mathbf{w})\,\mathbf{dx},\quad\forall\mathbf{w}\in\mathbf{V_0},$$

Now $\mathbf{y}$ is a tensor-valued function and $\mathbf{y}\boldsymbol{\nu} = \mathbf{y_{ij}}\boldsymbol{\nu}_\mathbf{j}$ is a vector–function defined on $\partial\Omega$.
If

$$\mathbf{y} \in \mathbf{Q}^* := \{\mathbf{y} \in \mathbf{Y}^* \mid \mathbf{divy} \in \mathbf{L}^2(\boldsymbol{\Omega}, \mathbb{M}^{\mathbf{n}\times\mathbf{n}}),\ \mathbf{y}\boldsymbol{\nu} \in \mathbf{L}^2(\partial_2\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{n}})\}.$$

then

$$\langle \boldsymbol{\Lambda}^*\mathbf{y}, \mathbf{w}\rangle = -\int_{\boldsymbol{\Omega}} \mathbf{div}\,\mathbf{y}\cdot\mathbf{w}\,\mathbf{dx} + \int\limits_{\partial_2\boldsymbol{\Omega}} (\mathbf{y}\boldsymbol{\nu})\cdot\mathbf{w}\,\mathbf{d}\boldsymbol{\Gamma}$$

so that

$$\boldsymbol{\Lambda}^*\mathbf{y} = \{-\mathbf{divy}\mid_{\boldsymbol{\Omega}},\quad (\mathbf{y}\boldsymbol{\nu})\mid_{\partial_2\boldsymbol{\Omega}}\}.$$

Upper estimates

By applying the general estimate, we obtain the following upper estimate:

$$\frac{1}{2}\int_\Omega \mathbb{L}\,\varepsilon(\mathbf{v}-\mathbf{u}) : \varepsilon(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \leq \mathbf{M}_\oplus(\mathbf{v},\beta,\mathbf{y}),$$

where

$$\mathbf{M}_\oplus(\mathbf{v},\beta,\mathbf{y}) = \frac{1+\beta}{2}\mathbf{D}(\varepsilon\mathbf{v},\mathbf{y}) + \frac{1+\beta}{2\beta}[\![\,\Lambda^*\mathbf{y}+\ell\,]\!]^2$$

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

$$[\![\boldsymbol{\Lambda}^*\mathbf{y} + \ell]\!] = \sup_{\mathbf{w} \in \mathbf{V}_0} \frac{\langle \boldsymbol{\Lambda}^*\mathbf{y} + \ell, \mathbf{w} \rangle}{|\!|\!| \boldsymbol{\Lambda}\mathbf{w} |\!|\!|} =$$

$$= \sup_{\mathbf{w} \in \mathbf{V}_0} \frac{\int_\Omega (\mathbf{y} : \varepsilon(\mathbf{w}) + \mathbf{f} \cdot \mathbf{w})\, \mathbf{dx} + \int\limits_{\partial_2\Omega} \mathbf{F} \cdot \mathbf{w}\, \mathbf{ds}}{|\!|\!| \varepsilon(\mathbf{w}) |\!|\!|} =$$

$$= \sup_{\mathbf{w} \in \mathbf{V}_0} \frac{\int_\Omega (\mathbf{f} - \mathbf{div}\, \mathbf{y}) \cdot \mathbf{w}\, \mathbf{dx} + \int\limits_{\partial_2\Omega} (\mathbf{F} + \mathbf{y}\nu) \cdot \mathbf{w}\, \mathbf{ds}}{|\!|\!| \varepsilon(\mathbf{w}) |\!|\!|} \leq$$

$$\leq \sup_{\mathbf{w} \in \mathbf{V}_0} \frac{\|\mathbf{f} - \mathbf{div}\, \mathbf{y}\|_{2,\Omega} \|\mathbf{w}\|_{2,\Omega} + \|\mathbf{F} + \mathbf{y}\nu\|_{\partial_2\Omega} \|\mathbf{w}\|_{\partial_2\Omega}}{|\!|\!| \varepsilon(\mathbf{w}) |\!|\!|}.$$

Let $\mathbf{C}_\Omega$ be a constant in the inequality

$$\int_\Omega |\mathbf{w}|^2 \, \mathbf{dx} + \int_{\partial_2\Omega} |\mathbf{w}|^2 \, \mathbf{ds} \leq \mathbf{C}_\Omega^2 \|\varepsilon(\mathbf{w})\|_\Omega^2, \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

Note that the existence of such a constant follows from the Korn's inequality. Indeed, the inequality

$$\int_\Omega |\mathbf{w}|^2 \, \mathbf{dx} + \int_{\partial_2\Omega} |\mathbf{w}|^2 \, \mathbf{ds} \leq \hat{\mathbf{C}}_\Omega^2 \|\nabla(\mathbf{w})\|_\Omega^2, \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

for the tensor–gradient $\nabla(\mathbf{w})$ follows from the Friederichs type inquality for the vector–valued functions and the respective trace theorems. By (6.15) we recall that for the functions in $\mathbf{V_0}$

$$\|\nabla(\mathbf{w})\|_\Omega \leq \mathbf{C}\|\varepsilon(\mathbf{w})\|_\Omega$$

with a certain constant $\mathbf{C}$ and the estimate follows.

In practice, values of $\mathbf{C}_{\boldsymbol{\Omega}}$ can be estimated by minimizing the quotient

$$\frac{\|\varepsilon(\mathbf{w})\|^2_{\boldsymbol{\Omega}}}{\int_\Omega |\mathbf{w}|^2\,\mathbf{dx} + \int\limits_{\partial_2\boldsymbol{\Omega}} |\mathbf{w}|^2\,\mathbf{ds}}$$

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 $[\![\, \boldsymbol{\Lambda}^*\mathbf{y} + \ell \,]\!]$.
By the inequality $ab + cd \leq \sqrt{a^2 + c^2}\sqrt{b^2 + d^2}$, we obtain

$$[\![\, \boldsymbol{\Lambda}^*\mathbf{y} + \ell \,]\!] \leq$$

$$\leq \left(\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|^2_{\boldsymbol{\Omega}} + \|\mathbf{F} + \mathbf{y}\boldsymbol{\nu}\|^2_{\partial_2\boldsymbol{\Omega}}\right)^{1/2} \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{(\|\mathbf{w}\|^2_{\boldsymbol{\Omega}} + \|\mathbf{w}\|^2_{\partial_2\boldsymbol{\Omega}})^{\mathbf{1/2}}}{|\!|\!|\ \varepsilon(\mathbf{w})\ |\!|\!|} \leq$$

$$\leq \mathbf{C}_{\boldsymbol{\Omega}}\mathbf{c}_{\mathbf{1}}^{-\mathbf{1/2}} \left(\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|^2_{\boldsymbol{\Omega}} + \|\mathbf{F} + \mathbf{y}\boldsymbol{\nu}\|^2_{\partial_2\boldsymbol{\Omega}}\right)^{\mathbf{1/2}}$$

Error Majorant for mixed boundary conditions

Hence, we arrive at the Majorant $\mathbf{M}_{\oplus}$:

$$\mathbf{M}_{\oplus}(\varepsilon(\mathbf{v}),\mathbf{y}) = \frac{\mathbf{1}+\beta}{\mathbf{2}}\int_{\Omega}(\varepsilon(\mathbf{u})-\mathbb{L}^{-\mathbf{1}}\mathbf{y}):(\mathbb{L}\,\varepsilon(\mathbf{u})-\mathbf{y})\,\mathbf{dx}+ \\ +\frac{1+\beta}{2\beta c_1}\,\mathbf{C}_{\Omega}^{2}\left(\|\mathbf{div}\,\mathbf{y}-\mathbf{f}\|_{\Omega}^{2}+\|\mathbf{F}+\mathbf{y}\nu\|_{\partial_2\Omega}^{2}\right). \quad (6.17)$$

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 $\{\mathbf{Y}^*_{\mathbf{k}}\} \subset \mathbf{H}^1(\boldsymbol{\Omega}, \mathbb{M}^{\mathbf{n}\times\mathbf{n}})$ be a collection of finite-dimensional subspaces that satisfy the limit density condition. Then, (6.17) generates a sequence of computable upper bounds

$$\mathbf{M}^{\mathbf{k}}_{\oplus} = \inf_{\substack{\mathbf{y}\in\mathbf{Y}^*_{\mathbf{k}} \\ \boldsymbol{\beta}\in\mathbb{R}_+}} \left\{ \frac{1+\beta}{2}\int_{\Omega}\left(\mathbb{L}\,\varepsilon(\mathbf{v}) : \varepsilon(\mathbf{v}) + \mathbb{L}^{-\mathbf{1}}\mathbf{y} : \mathbf{y} - \mathbf{2}\varepsilon(\mathbf{v}) : \mathbf{y}\right) dx \right.$$
$$\left. + \frac{1+\beta}{2\beta c_1}\,\mathbf{C}^2_{\boldsymbol{\Omega}}\left(\|\mathbf{div}\,\mathbf{y} - \mathbf{f}\|^2_{\boldsymbol{\Omega}} + \|\mathbf{F} + \mathbf{y}\nu\|^2_{\partial_2\boldsymbol{\Omega}} \right)\right\},$$

which tends to the exact value of the error.

Lower estimates

Lower estimates also follow from the general theory. We have

$$\frac{1}{2}\int_\Omega \mathbb{L}\,\varepsilon(\mathbf{v}-\mathbf{u}):\varepsilon(\mathbf{v}-\mathbf{u})\,\mathbf{dx} \geq \mathbf{M}_\ominus(\mathbf{v},\mathbf{w}), \quad \forall \mathbf{w}\in\mathbf{V_0},$$

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 $\frac{1}{2}|\!|\!|\varepsilon(\mathbf{v}-\mathbf{u})|\!|\!|^{\mathbf{2}}$.

FUNCTIONAL A POSTERIORI ESTIMATES. FOURTH ORDER EQUATIONS.

Linear elliptic equations of the fourth order

Now, we consider the problem

$$\nabla \cdot \nabla \cdot (\mathbf{B}\nabla\nabla \mathbf{u}) = \mathbf{f} \quad \text{in} \quad \boldsymbol{\Omega}, \tag{6.18}$$

$$\mathbf{u} = \frac{\partial \mathbf{u}}{\partial \boldsymbol{\nu}} = \mathbf{0} \quad \text{on} \quad \partial\boldsymbol{\Omega}. \tag{6.19}$$

Here $\Omega \subset \mathbb{R}^2$, ν denotes the outward unit normal to the boundary, and $\mathbf{B} = \{\mathbf{b_{ijkl}}\} \in \mathcal{L}(\mathbb{M}_{\mathbf{s}}^{\mathbf{2}\times\mathbf{2}}, \mathbb{M}_{\mathbf{s}}^{\mathbf{2}\times\mathbf{2}})$. We assume that $\mathbf{b_{ijkl}} = \mathbf{b_{jikl}} = \mathbf{b_{klij}}$,

$$\alpha_{\mathbf{1}}|\boldsymbol{\eta}|^{\mathbf{2}} \leq \mathbf{B}\boldsymbol{\eta} : \boldsymbol{\eta} \leq \alpha_{\mathbf{2}}|\boldsymbol{\eta}|^{\mathbf{2}}, \quad \forall \boldsymbol{\eta} \in \mathbb{M}_{\mathbf{s}}^{\mathbf{2}\times\mathbf{2}},$$

and

$$\mathbf{f} \in \mathbf{L}^{\mathbf{2}}(\boldsymbol{\Omega}), \quad \mathbf{b_{ijkl}} \in \mathbf{L}^{\infty}(\boldsymbol{\Omega}).$$

To apply the general scheme, we set

$$\mathbf{U} = \mathbf{L}^2(\mathbf{\Omega}, \mathbb{M}_s^{2\times 2}), \qquad \mathbf{V} = \mathbf{H}^2(\mathbf{\Omega}),$$
$$\mathbf{V_0} = \{\mathbf{w} \in \mathbf{V} \mid \mathbf{w} = \frac{\partial \mathbf{w}}{\partial \nu} = \mathbf{0} \ \text{ on } \partial\mathbf{\Omega}\},$$

and define $\mathbf{\Lambda}$ as the Hessian operator. Now, the basic integral identity has the form

$$\int_{\mathbf{\Omega}} \mathbf{B}\nabla\nabla\mathbf{u} : \nabla\nabla\mathbf{w}\,\mathbf{dx} = \int_{\mathbf{\Omega}} \mathbf{fw}\,\mathbf{dx} \quad \forall \mathbf{w} \in \mathbf{V_0}. \tag{6.20}$$

By $\mathbf{B}^{-1}$ we denote the inverse tensor, which satisfies the double inequality

$$\alpha_2^{-1}|\eta|^2 \le \mathbf{B}^{-1}\eta : \eta \le \alpha_1^{-1}|\eta|^2, \quad \forall \eta \in \mathbb{M}_s^{2\times 2},$$

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

$$\mathbf{Q}_\ell^* = \{\mathbf{y} \in \mathbf{Y}^* \mid \int_\Omega \mathbf{y} : \nabla\nabla\mathbf{w}\, \mathbf{dx} = \int_\Omega \mathbf{f}\mathbf{w}\, \mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{V_0}\}.$$

Since

$$\|\nabla\nabla\mathbf{w}\| \geq \alpha_3 \|\mathbf{w}\|_{2,2,\Omega} \quad \forall \mathbf{w} \in \mathbf{V_0},$$

we have the required version of the coercivity condition

$$\|\mathbf{\Lambda w}\| \geq \mathbf{c_3} \|\mathbf{w}\|_{\mathbf{V}}.$$

Problem (6.18) and (6.19) 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

$$\mathbf{I}^*(\mathbf{q}) = -\frac{1}{2} \int_{\Omega} \mathbf{B}^{-1} \mathbf{q} : \mathbf{q} \, \mathbf{dx}.$$

By In this case, the two basic relations for deviations derived in Lecture 5 come in the form:

$$\| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 + \| \mathbf{q}-\mathbf{p} \|_*^2 = 2(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q})), \qquad (6.21)$$

and

$$\| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 + \| \mathbf{q}-\mathbf{p} \|_*^2 = 2\mathbf{D}(\nabla\nabla\mathbf{v},\mathbf{q}) =$$
$$= \int_\Omega \left(\mathbf{B}\nabla\nabla\mathbf{v} : \nabla\nabla\mathbf{v} + \mathbf{B}^{-1}\mathbf{q} : \mathbf{q} - 2\nabla\nabla\mathbf{v} : \mathbf{q} \right) \mathbf{dx}, \qquad (6.22)$$

which 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:

$$\frac{1}{2} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 \leq (1+\beta)\mathbf{D}(\nabla\nabla\mathbf{v},\mathbf{y}) + \left(1+\frac{1}{\beta}\right)\frac{\mathbf{d}_\ell^2(\mathbf{y})}{2}, \qquad (6.23)$$

where $\mathbf{d}_\ell^2(\mathbf{y}) = \inf\limits_{\mathbf{q}\in\mathbf{Q}_\ell^*} \| \mathbf{q}-\mathbf{y} \|_*^2$.

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,

$$\mathbf{d}^2_\ell(\mathbf{y}) = [\![\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,]\!] = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{y} : \nabla\nabla \mathbf{w} - \mathbf{fw})\, \mathbf{dx}}{\|\!| \nabla\nabla \mathbf{w} \|\!|}.$$

If

$$\mathbf{y} \in \mathbf{H}(\mathbf{divdiv}, \boldsymbol{\Omega}) := \left\{ \mathbf{y} \in \mathbf{L}^2(\boldsymbol{\Omega}, \mathbb{M}^{\mathbf{n}\times\mathbf{n}}_{\mathbf{s}}) \mid \mathbf{divdiv}\,\mathbf{y} \in \mathbf{L}^2(\boldsymbol{\Omega}) \right\},$$

then this quantity is estimated by the relation

$$[\![\, \ell + \boldsymbol{\Lambda}^*\mathbf{y} \,]\!] \leq \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\|\mathbf{divdiv}\,\mathbf{y} - \mathbf{f}\|_{\boldsymbol{\Omega}} \|\mathbf{w}\|_{\boldsymbol{\Omega}}}{|\!|\!| \nabla\nabla\mathbf{w} |\!|\!|} \leq$$

$$\leq \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\|\mathbf{divdiv}\,\mathbf{y} - \mathbf{f}\|_{\boldsymbol{\Omega}} \|\mathbf{w}\|_{\boldsymbol{\Omega}}}{\alpha_1 \|\nabla\nabla\mathbf{w}\|} \leq \frac{\mathbf{C_{1\Omega}}}{\alpha_1} \|\mathbf{divdiv}\,\mathbf{y} - \mathbf{f}\|_{\boldsymbol{\Omega}},$$

in which $\mathbf{C_{1\Omega}}$ is a constant in the inequality

$$\|\mathbf{w}\|_{\boldsymbol{\Omega}} \leq \mathbf{C_{1\Omega}} \|\nabla\nabla\mathbf{w}\|_{\boldsymbol{\Omega}} \qquad \forall \mathbf{w} \in \mathbf{V_0}.$$

Now, we obtain the first variant of a posteriori estimate for the biharmonic type problem.

First a posteriori estimate

$$\frac{1}{2} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 \leq (1+\beta)\mathbf{D}(\nabla\nabla\mathbf{v},\mathbf{y}) + \\ + \left(1+\frac{1}{\beta}\right)\frac{\mathbf{C}^2_{1\Omega}}{2\alpha_1^2}\|\mathbf{divdiv}\,\mathbf{y}-\mathbf{f}\|^2_{\Omega}, \quad (6.24)$$

Here, $\mathbf{y}$ is an arbitrary tensor-valued function from $\mathbf{H}(\mathbf{div}\,\mathbf{div},\mathbf{\Omega})$ 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.

$$
[\![\, \ell + \boldsymbol{\Lambda}^* \mathbf{y} \,]\!] = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{y} : \nabla\nabla \mathbf{w} - \mathbf{f}\mathbf{w})\, \mathbf{dx}}{\| \nabla\nabla \mathbf{w} \|} =
$$

$$
= \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_\Omega (\mathbf{y} : \nabla\nabla \mathbf{w} + \boldsymbol{\eta} \cdot \nabla \mathbf{w} + \mathbf{div}\boldsymbol{\eta}\mathbf{w} - \mathbf{f}\mathbf{w})\, \mathbf{dx}}{\| \nabla\nabla w \|} =
$$

$$
\frac{\int_\Omega (-\mathbf{divy} \cdot \nabla \mathbf{w} + \boldsymbol{\eta} \cdot \nabla \mathbf{w} + \mathbf{div}\boldsymbol{\eta}\mathbf{w} - \mathbf{f}\mathbf{w})\, \mathbf{dx}}{\| \nabla\nabla w \|} \leq
$$

$$
\leq \frac{\mathbf{C_{2\Omega}}}{\alpha_1} \| \mathbf{div}\, \mathbf{y} - \boldsymbol{\eta} \|_\Omega + \frac{\mathbf{C_{1\Omega}}}{\alpha_1} \| \mathbf{div}\, \boldsymbol{\eta} - \mathbf{f} \|_\Omega .
$$

Here, $\boldsymbol{\eta}$ is an arbitrary vector-valued function from $\mathbf{H}(\mathbf{div}, \boldsymbol{\Omega})$ 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

$$\frac{\mathbf{1}}{\mathbf{2}} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^{\mathbf{2}} \leq (\mathbf{1}+\beta)\mathbf{D}(\nabla\nabla\mathbf{v},\mathbf{y})+$$
$$+\left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\frac{\mathbf{1}}{\mathbf{2}\alpha_{\mathbf{1}}^{\mathbf{2}}}\left(\mathbf{C}_{\mathbf{2}\Omega}\|\mathbf{div}\,\mathbf{y}-\boldsymbol{\eta}\|_{\Omega}+\mathbf{C}_{\mathbf{1}\Omega}\|\mathbf{div}\,\boldsymbol{\eta}-\mathbf{f}\|_{\Omega}\right)^{\mathbf{2}}, \quad (6.25)$$

in which $\mathbf{y} \in \mathbf{\Sigma}_{\mathbf{div}}(\mathbf{\Omega})$ and $\boldsymbol{\eta} \in \mathbf{H}(\mathbf{div},\mathbf{\Omega})$.

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:

$$\frac{1}{2} \| \nabla\nabla(\mathbf{v}-\mathbf{u}) \|^2 \leq (1+\beta)\mathbf{D}(\nabla\nabla\mathbf{v},\mathbf{y}) + \\ + \left(1+\frac{1}{\beta}\right)\frac{\mathbf{C}^2_{2\Omega}}{2\alpha_1^2}\left(\|\mathbf{div}\,\mathbf{y}-\eta\|_\Omega + \mathbf{C_F}\|\mathbf{div}\,\eta - \mathbf{f}\|_\Omega\right)^2, \quad (6.26)$$

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

$$\frac{1}{2} \parallel\!| \nabla\nabla(\mathbf{v}-\mathbf{w}) \parallel\!|^2 \geq \mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w}) \quad \mathbf{w} \in \mathbf{V_0}, \tag{6.27}$$

where

$$\mathbf{M}_{\ominus}(\mathbf{v},\mathbf{w}) := -\frac{1}{2} \parallel\!| \nabla\nabla\mathbf{w} \parallel\!|^2 - \int_{\Omega} (\mathbf{B}\nabla\nabla\mathbf{v} : \nabla\nabla\mathbf{w} - \mathbf{f}\mathbf{w})\mathbf{dx}.$$

Lecture 7.
FUNCTIONAL A POSTERIORI ESTIMATES. STOKES PROBLEM.

Lecture plan

- **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 div$\mathbf{v} = \phi$;**
- **A posteriori estimates for problems on a subspace.**

Stokes problem

George Stokes

Classical formulation of the Stokes problem: find a vector–valued function **u** (velocity) and a scalar–valued function **p** (pressure) that satisfy the relations

$$-\nu\boldsymbol{\Delta}\mathbf{u} = \mathbf{f} - \nabla\mathbf{p} \qquad \text{in } \Omega, \tag{7.1}$$

$$\mathbf{div}\mathbf{u} = \mathbf{0} \qquad \text{in } \Omega, \tag{7.2}$$

$$\mathbf{u} = \mathbf{u_0} \qquad \text{on } \partial\Omega, \tag{7.3}$$

Nomenclature

Let smooth solenoidal functions with compact supports in Ω form the set be denoted by $\dot{J}^\infty(\Omega)$. The closure of $\dot{J}^\infty(\Omega)$ with respect to the norm $\|\nabla \mathbf{v}\|$ is the space $\overset{\circ}{\mathbf{J}}{}^1_2(\mathbf{\Omega})$.
Next, $\mathbf{W} := \mathbf{W}^1_2(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}})$ and $\mathbf{\Sigma} := \mathbf{L_2}(\mathbf{\Omega}, \mathbb{M}^{\mathbf{d}\times\mathbf{d}})$, where $\mathbb{M}^{\mathbf{d}\times\mathbf{d}}$ 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, $\overset{\circ}{\mathbf{J}}{}^1_2(\mathbf{\Omega}) + \mathbf{u_0}$ contains functions of the form $\mathbf{w} + \mathbf{u_0}, \mathbf{w} \in \overset{\circ}{\mathbf{J}}{}^1_2(\mathbf{\Omega})$.
The operator $\varepsilon(\mathbf{v}) := \frac{1}{2}(\nabla \mathbf{v} + (\nabla \mathbf{v})^{\mathbf{T}})$ acts from $\mathbf{W}$ to $\mathbf{\Sigma}$.

We will also use the Hilbert space $\boldsymbol{\Sigma}_{\rm div}(\boldsymbol{\Omega})$, which is a subspace of $\boldsymbol{\Sigma}$ 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

$$(\boldsymbol{\tau}, \boldsymbol{\eta}) := \int_{\boldsymbol{\Omega}} (\boldsymbol{\tau} : \boldsymbol{\eta} + \mathbf{div}\boldsymbol{\tau} \cdot \mathbf{div}\boldsymbol{\eta})\, \mathbf{dx}\,.$$

By $\overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\boldsymbol{\Omega})$ we denote the space of square summable functions with zero mean. Henceforth, we assume that

$$\mathbf{f} \in \mathbf{L_2}(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{d}}), \quad \mathbf{u_0} \in \mathbf{W^1_2}(\boldsymbol{\Omega}, \mathbb{R}^{\mathbf{d}}),$$

Generalized solution can be defined by the **integral identity**. It is a function $\mathbf{u} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}$ that meets the relation

$$\int_{\Omega} \nu \nabla(\mathbf{u}) : \nabla(\mathbf{v})\, \mathbf{dx} = \int_{\Omega} \mathbf{f} \cdot \mathbf{v}\, \mathbf{dx} \quad \forall \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega). \qquad (7.4)$$

It is well known that $\mathbf{u}$ exists and unique and can be viewed as the minimizer of the functional

$$\mathbf{I}(\mathbf{v}) = \int\limits_{\Omega} \left(\frac{\nu}{2} |\nabla(\mathbf{v})|^2 - \mathbf{f} \cdot \mathbf{v} \right) \mathbf{dx}$$

on the set $\overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}$. Thus, the problem

$$\inf_{\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\Omega) + \mathbf{u_0}} \mathbf{I}(\mathbf{v})$$

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 $\mathbf{L} : (\mathbf{W_0} + \mathbf{u_0}) \times \overset{\circ}{\mathbf{L}}_2(\boldsymbol{\Omega}) \to \mathbb{R}$ be defined as follows:

$$\mathbf{L}(\mathbf{v}, \mathbf{q}) = \int_{\boldsymbol{\Omega}} \left(\frac{\boldsymbol{\nu}}{\mathbf{2}} |\nabla \mathbf{v}|^{\mathbf{2}} - \mathbf{f} \cdot \mathbf{v} - \mathbf{q}\,\mathrm{\mathbf{div}}\mathbf{v} \right) \mathbf{dx}.$$

Now, **u** and **p** are defined as a saddle–point that satisfies the relations

$$\mathbf{L}(\mathbf{u}, \mathbf{q}) \leq \mathbf{L}(\mathbf{u}, \mathbf{p}) \leq \mathbf{L}(\mathbf{v}, \mathbf{p}) \qquad \forall \mathbf{v} \in \mathbf{W_0} + \mathbf{u_0},\ \mathbf{p} \in \overset{\circ}{\mathbf{L}}_2(\boldsymbol{\Omega}).$$

Extension of solenoidal fields and related results

Olga Ladyzhenskaya

First, we recall some basic results that has been established when the solvability of the Stokes problem was investigated. Works of O.A. Ladyzhenskaya made a grate 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

$$\|\nabla \bar{\mathbf{u}}\| \leq \boldsymbol{\kappa}_1(\boldsymbol{\Omega}) \|\mathbf{a}\|_{1/2,\partial\Omega}, \tag{7.5}$$

where $\kappa_1(\Omega)$ 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 $\mathbf{f} \in \overset{\circ}{\mathbf{L}}_2(\Omega)$ there exists a function $\bar{\mathbf{u}} \in \mathbf{W_0}$ satisfying the relation $\mathbf{div}\bar{\mathbf{u}} = \mathbf{f}$ and the condition

$$\|\nabla\bar{\mathbf{u}}\| \leq \kappa_2(\Omega)\|\mathbf{f}\|, \tag{7.6}$$

where $\kappa_2(\Omega)$ 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 $\mathbb{C}_\Omega$ such that**

$$\inf_{\substack{\phi\in\overset{\circ}{L}_2(\Omega)\\ \phi\neq 0}} \sup_{\substack{\mathbf{w}\in\mathbf{W}_0\\ \mathbf{w}\neq\mathbf{0}}} \frac{\int_\Omega \phi\, \mathbf{div}\mathbf{w}\, \mathbf{dx}}{\|\phi\|\;\|\nabla\mathbf{w}\|} \geq \mathbb{C}_\Omega\,. \tag{7.7}$$

Ivo Babuška

Franco Brezzi

Inf-Sup condition (7.7) 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 $\phi \in \overset{\circ}{\mathbf{L}}_2(\Omega)$ has a counterpart function $\mathbf{v}_\phi \in \mathbf{W}_0$ that meets the conditions

$$\mathbf{div}\mathbf{v}_\phi = \phi, \quad \|\nabla \mathbf{v}_\phi\| \leq \boldsymbol{\kappa}_2(\boldsymbol{\Omega})\|\phi\|.$$

In this case,

$$\sup_{\mathbf{v}\in\mathbf{W}_0,\mathbf{w}\neq 0} \frac{\int_\Omega \phi \,\mathbf{div}\mathbf{v}\,\mathbf{dx}}{\|\nabla \mathbf{v}\|\,\|\phi\|} \geq \frac{\int_\Omega \phi\,\mathbf{div}\mathbf{v}_\phi\,\mathbf{dx}}{\|\nabla \mathbf{v}_\phi\|\,\|\phi\|} = \frac{\|\phi\|}{\|\nabla \mathbf{v}_\phi\|} \geq \frac{\mathbf{1}}{\boldsymbol{\kappa}_2(\boldsymbol{\Omega})}$$

and, consequently, Inf-Sup condition holds with

$$\boxed{\mathbb{C}_\Omega = \frac{1}{\kappa_2(\Omega)}.}$$

It is easy to observe that the Inf-Sup condition can be presented in the form

$$\sup_{\substack{\mathbf{w}\in\mathbf{W_0}\\ \mathbf{w}\neq\mathbf{0}}} \frac{\displaystyle\int_{\Omega} \mathbf{p}\,\mathbf{div}\mathbf{w}\,\mathbf{dx}}{\|\nabla\mathbf{w}\|} \geq \mathbb{C}_{\Omega}\,\|\,\mathbf{p}\| \quad \text{for all } \mathbf{p}\in \overset{\circ}{\mathbf{L}}_2(\Omega).$$

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

$$[\![\,\nabla\mathbf{p}\,]\!] := \sup_{\mathbf{w}\in\mathbf{W_0}} \frac{<\nabla\mathbf{p},\mathbf{w}>}{\|\nabla\mathbf{w}\|}.$$

Then, we arrive to the Nečas inequality.

Nečas inequality

Jindřich Nečas

$$\|\mathbf{p}\| \leq \kappa_2 \, [\![\nabla p]\!] \qquad \forall \ \mathbf{p} \in \overset{\circ}{\mathbf{L}}_2(\Omega), \tag{7.8}$$

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 $\mathbb{C}_\Omega$ 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 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 q is compact.)

Since

$$\mathbf{J}(\mathbf{v}) = \sup_{\mathbf{q}\in\boldsymbol{\Sigma}} \mathbf{L}(\mathbf{v},\mathbf{q}) \geq \mathbf{L}(\bar{\mathbf{q}},\mathbf{v}),$$

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 $\mathbf{q} \in \overset{\circ}{\mathbf{L}}_2(\boldsymbol{\Omega})$ the mapping

$$\mathbf{v} \mapsto \mathbf{L}(\mathbf{v}, \mathbf{q}) = \int_{\boldsymbol{\Omega}} \left(\frac{\boldsymbol{\nu}}{\mathbf{2}} |\nabla \mathbf{v}|^{\mathbf{2}} - \mathbf{f} \cdot \mathbf{v} - \mathbf{q} \mathbf{div} \mathbf{v} \right) \mathbf{dx}.$$

is convex and continuous (in $\mathbf{W}$) and there exists am element $\bar{\mathbf{q}} \in \overset{\circ}{L}_2(\Omega)$ (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 $\overset{\circ}{\mathbf{L}}_2(\Omega)$) for any $\mathbf{v} \in \mathbf{V}$. Therefore, existence of a saddle point is guaranteed provided that the coercivity condition

$$\lim_{\|\mathbf{q}\| \to +\infty} \inf_{\mathbf{v} \in \mathbf{W_0} + \mathbf{u_0}} \mathbf{L}(\mathbf{v}, \mathbf{q}) = -\infty \qquad (7.9)$$

is established. By Lemma 2 we can prove this fact.

Consider the functional

$$\mathbf{l}(\mathbf{q}) := \inf_{\mathbf{v}\in\mathbf{W_0}+\mathbf{u_0}} \mathbf{L}(\mathbf{v},\mathbf{q})$$

and the variational problem

$$\mathbf{l}(\mathbf{p}) = \sup_{\mathbf{q}\in\overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\boldsymbol{\Omega})} \mathbf{l}(\mathbf{q}) \tag{7.10}$$

for the **pressure** function. Note that the functional $\mathbf{l}$ has no explicit integral-type form and is defined as a supremum–functional. The solvability of this problem follows from the coercivity condition (7.9). To prove (7.9) we apply Lemma 2.

Coercivity of the variational problem for the pressure function

Indeed, by Lemma 2 for any $\mathbf{q} \in \overset{\circ}{\mathbf{L}}_2(\boldsymbol{\Omega})$ we find $\mathbf{v_q} \in \mathbf{W_0}$ such that

$$\mathbf{div v_q} = \mathbf{q} \quad \text{and} \quad \|\nabla \mathbf{v_q}\| \leq \boldsymbol{\kappa}_2 \|\mathbf{q}\|.$$

Take $\mathbf{v} = \mu \mathbf{v_q} + \mathbf{u_0}$ and recall that $\mathbf{div u_0} = \mathbf{0}$. Then,

$$\inf_{\mathbf{v} \in \mathbf{W_0} + \mathbf{u_0}} \mathbf{L}(\mathbf{v}, \mathbf{q}) \leq \int_\Omega \Big(\frac{\boldsymbol{\nu}}{\mathbf{2}} |\nabla(\mu \mathbf{v_q} + \mathbf{u_0})|^2 - \mathbf{f} \cdot (\mu \mathbf{v_q} + \mathbf{u_0}) - \mathbf{q div}(\mu \mathbf{v_q} + \mathbf{u_0}) \Big) \mathbf{dx} \leq$$

$$\leq \int_\Omega \left(\frac{\boldsymbol{\nu}}{\mathbf{2}} |\nabla \mathbf{u_0}|^2 - \mathbf{f} \cdot \mathbf{u_0} \right) \mathbf{dx} + \mu(\boldsymbol{\nu} \|\nabla \mathbf{u_0}\| + \mathbf{C_\Omega} \|\mathbf{f}\|) \|\nabla \mathbf{v_q}\| +$$

$$+ \frac{\boldsymbol{\nu} \mu^2}{\mathbf{2}} \|\nabla \mathbf{v_q}\|^2 - \mu \|\mathbf{q}\|^2 \leq \int_\Omega \left(\frac{\boldsymbol{\nu}}{\mathbf{2}} |\nabla \mathbf{u_0}|^2 - \mathbf{f} \cdot \mathbf{u_0} \right) \mathbf{dx} +$$

$$+ \mu(\boldsymbol{\nu} \|\nabla \mathbf{u_0}\| + \mathbf{C_\Omega} \|\mathbf{f}\|) \boldsymbol{\kappa}_2 \|\mathbf{q}\| + \mu \left(\frac{\boldsymbol{\nu} \mu \boldsymbol{\kappa}_2^2}{\mathbf{2}} - \mathbf{1} \right) \|\mathbf{q}\|^2,$$

where $\mathbf{C_\Omega}$ is a constant in the Friederichs inequality.

We see that

$$\mathbf{l}(\mathbf{q}) \leq \quad c_1(\mathbf{u_0}, \mathbf{f}, \boldsymbol{\nu}) + \mu(\boldsymbol{\nu}\|\nabla\mathbf{u_0}\| + \mathbf{C_\Omega}\|\mathbf{f}\|)\boldsymbol{\kappa_2}\|\mathbf{q}\| + \\ +\mu\left(\frac{\boldsymbol{\nu}\mu\boldsymbol{\kappa_2^2}}{\mathbf{2}} - \mathbf{1}\right)\|\mathbf{q}\|^{\mathbf{2}}.$$

Set here $\mu = \frac{1}{\boldsymbol{\nu}\boldsymbol{\kappa}_2^2}$. Then

$$\inf_{\mathbf{v}\in\mathbf{W_0}+\mathbf{u_0}} \mathbf{L}(\mathbf{v}, \mathbf{q}) \quad \leq c_1 + c_2\|\mathbf{q}\| - \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\nu}\boldsymbol{\kappa_2^2}}\|\mathbf{q}\|^{\mathbf{2}} \to -\infty \quad \text{as } \|\mathbf{q}\| \to +\infty.$$

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 $\widehat{v} \in H^1$ and the space of solenoidal functions**.

Estimates in L_2-norm. An estimate of the distance between $\widehat{\mathbf{v}}$ and the space

$$\mathbf{J}_2^1(\Omega) := \left\{ \mathbf{v} \in \mathbf{W}_2^1(\Omega) \mid \mathbf{div v} = \mathbf{0} \right\}$$

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 $\phi \in \overset{\circ}{\mathbf{W}}{}_2^1(\mathbf{\Omega})$ for any $\mathbf{f} \in \mathbf{L_2}(\mathbf{\Omega})$, we conclude that for any $\mathbf{f}$ there exists $\mathbf{v_f} = \nabla\phi$ such that

$$\mathbf{div v_f} = \mathbf{f} \qquad \text{and} \qquad \|\mathbf{v_f}\| \leq \mathbf{C_F}\|\mathbf{f}\|.$$

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 $\mathbf{J}_2^1$ and we observe that

$$\|\widehat{\mathbf{v}} - \mathbf{v_0}\| \leq \mathbf{C_F}\|\mathbf{div}\widehat{\mathbf{v}}\|.$$

Estimates in $\mathbf{H^1}$-norm. Let now $\widehat{\mathbf{v}} \in \mathbf{W_0}$. Set $\mathbf{f} = \mathbf{div}\widehat{\mathbf{v}}$. Since

$$\int_{\Omega} \mathbf{div}\widehat{\mathbf{v}}\, \mathbf{dx} = \int_{\partial\Omega} \mathbf{v} \cdot \nu\, \mathbf{ds} = \mathbf{0},$$

we see that $\mathbf{f} \in \overset{\circ}{\mathbf{L}}_{\mathbf{2}}(\Omega)$. Then, by Lemma 2, one can find $\mathbf{u_f} \in \mathbf{W_0}$ such that

$$\mathbf{divu_f} = \mathbf{div}\widehat{\mathbf{v}}, \quad \text{and} \quad \|\nabla \mathbf{u_f}\| \leq \kappa_2(\Omega)\|\mathbf{div}\widehat{\mathbf{v}}\|.$$

In other words, there exists a solenoidal field $\mathbf{w_0} = (\widehat{\mathbf{v}} - \mathbf{u_f}) \in \mathbf{W_0}$ such that

$$\|\nabla(\widehat{\mathbf{v}} - \mathbf{w_0})\| = \|\nabla \widehat{\mathbf{u}}_{\mathbf{f}}\| \leq \kappa_2(\Omega)\|\mathbf{div}\widehat{\mathbf{v}}\|.$$

This fact can be presented in another form

$$\inf_{\mathbf{v}\in \overset{\circ}{\mathbf{J}}{}^{1}_{2}(\boldsymbol{\Omega})} \|\nabla(\widehat{\mathbf{v}}-\mathbf{v})\| \leq \boldsymbol{\kappa}_{2}(\boldsymbol{\Omega})\|\mathbf{div}\widehat{\mathbf{v}}\|. \tag{7.11}$$

Thus, for the functions with zero traces the distance to $\overset{\circ}{\mathbf{J}}{}^{1}_{2}(\boldsymbol{\Omega})$ in a strong norm is also measured via $\|\mathbf{div}\widehat{\mathbf{v}}\|$, but with a different factor: $\boldsymbol{\kappa}_2(\Omega)$.

Comments on the value of $\mathbb{C}_\Omega$

Note that $\mathbb{C}_\Omega$ can be estimated throughout the constant $\mathbf{C_F}$ and the constant $\mathbf{C_P}$ in the Poincare inequality. Indeed,

$$\mathbb{C}_\Omega = \inf_{\mathbf{q} \in \overset{\circ}{\mathbf{L}}_2,\, \mathbf{q} \neq \mathbf{0}} \mathcal{E}(\mathbf{q}),$$

$$\mathcal{E}(\mathbf{q}) = \sup_{\mathbf{w} \in \mathbf{W_0},\, \mathbf{w} \neq \mathbf{0}} \frac{\int_\Omega \mathbf{q}\, \mathbf{div w}\, \mathbf{dx}}{\|\mathbf{q}\|\, \|\nabla \mathbf{w}\|}.$$

For $\mathbf{q} \in \overset{\sim}{\mathbf{W}} (\boldsymbol{\Omega}) := \overset{\circ}{\mathbf{L}}_2 \cap \mathbf{W}^1_2(\boldsymbol{\Omega})$ we have

$$\mathcal{E}(\mathbf{q}) = \sup_{\mathbf{w} \in \mathbf{W_0},\, \mathbf{w} \neq \mathbf{0}} \frac{\int_\Omega \nabla \mathbf{q} \cdot \mathbf{w}\, \mathbf{dx}}{\|\mathbf{q}\|\, \|\nabla \mathbf{w}\|} \leq \frac{\|\nabla \mathbf{q}\|}{\|\mathbf{q}\|} \sup_{\mathbf{w} \in \mathbf{W_0},\, \mathbf{w} \neq \mathbf{0}} \frac{\|\mathbf{w}\|}{\|\nabla \mathbf{w}\|}$$
$$\leq \mathbf{C_F} \frac{\|\nabla \mathbf{q}\|}{\|\mathbf{q}\|}.$$

Let $\mathbf{C_P}$ be the smallest constant in the inequality

$$\|\mathbf{q}\| \leq \mathbf{C_P}\|\nabla \mathbf{q}\|, \quad \mathbf{q} \in \widetilde{\mathbf{W}}(\mathbf{\Omega}),$$

i.e.,

$$\inf_{\mathbf{q}\in\widetilde{\mathbf{W}}(\mathbf{\Omega}),\, \mathbf{q}\neq\mathbf{0}} \frac{\|\nabla \mathbf{q}\|}{\|\mathbf{q}\|} = \frac{1}{\mathbf{C_P}}.$$

Then

$$\mathbb{C}_\Omega = \inf_{\mathbf{q}\in\overset{\circ}{\mathbf{L}}_2,\, \mathbf{q}\neq\mathbf{0}} \mathcal{E}(\mathbf{q}) \leq \inf_{\mathbf{q}\in\widetilde{\mathbf{W}}(\mathbf{\Omega}),\, \mathbf{q}\neq\mathbf{0}} \mathcal{E}(\mathbf{q}) \leq \frac{\mathbf{C_F}}{\mathbf{C_P}}.$$

LBB-condition can be written in the form

$$\|\mathbf{p}\| \leq \mathbb{C}_\Omega{}^{-1} \,[\![\nabla p]\!] \quad \forall\; p \in \overset{\circ}{\mathbf{L}}_2,$$

what amounts

$$\mathbb{C}_{\boldsymbol{\Omega}} \leq \frac{[\![\nabla \mathbf{p}]\!]}{\|\mathbf{p}\|}$$

we see the meaning of this constant: $\mathbb{C}_\Omega$ **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 $\boldsymbol{\Omega} \in \mathbb{R}^{\mathbf{n}}$ then

$$\frac{\|\nabla \mathbf{p}\|_{(-1)}}{\|\mathbf{p}\|} \leq \mathbf{n} \qquad \forall \mathbf{p} \in \mathbf{L}_2(\boldsymbol{\Omega}).$$

Proof.

$$\sup_{\substack{\mathbf{w} \in \mathbf{W_0} \\ \mathbf{w} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\, \mathbf{div w}\, \mathbf{dx}}{\|\nabla \mathbf{w}\|} =$$

$$\sup_{\substack{\mathbf{w} \in \mathbf{W_0} \\ \mathbf{w} \neq \mathbf{0}}} \frac{\sum_{\mathbf{t=1}}^{\mathbf{n}} \int_\Omega \mathbf{p}\, \mathbf{w_{t,t}}\, \mathbf{dx}}{\|\nabla \mathbf{w}\|} \leq \sum_{\mathbf{t=1}}^{\mathbf{n}} \sup_{\substack{\mathbf{w_t} \in \mathbf{W_0} \\ \mathbf{w_t} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\, \mathbf{w_{t,t}}\, \mathbf{dx}}{\|\nabla \mathbf{w}\|}.$$

Since

$$\|\nabla \mathbf{w}\|^2 = \int_\Omega \Big(\sum_{t,s=1,n}^{n} w_{t,s}^2 \Big)\, dx \geq \int_\Omega w_{t,t}^2\, dx \quad \forall t = 1, 2, \ldots n$$

we have

$$\sup_{\substack{\mathbf{w} \in \mathbf{W}_0 \\ \mathbf{w} \neq 0}} \frac{\displaystyle\int_\Omega p\, \mathrm{div} \mathbf{w}\, dx}{\|\nabla \mathbf{w}\|} \leq \sum_{t=1}^{n} \sup_{\substack{w_t \in W_0 \\ w_t \neq 0}} \frac{\displaystyle\int_\Omega p\, w_{t,t}\, dx}{\|w_{t,t}\|} \leq$$

$$\leq \sum_{t=1}^{n} \sup_{\substack{\eta \in L_2 \\ \eta \neq 0}} \frac{\displaystyle\int_\Omega p\, \eta\, dx}{\|\eta\|} = \sum_{t=1}^{n} \|p\| = n\, \|p\| .$$

Proposition 2

If $n = 1$ then $\mathbb{C}_\Omega = 1$.

Let $\boldsymbol{\Omega} = (\mathbf{a}, \mathbf{b})$. Due to Proposition 1 we see that $\mathbb{C}_\Omega \leq 1$. Let $\mathbf{p}$ be an arbitrary function from the set $\overset{\circ}{\mathbf{L}}_\mathbf{2}$. Then, the function

$$\mathbf{w}^{(\mathbf{p})} = \int\limits_{\mathbf{a}}^{\mathbf{x}} \mathbf{p}\,\mathbf{dx} \in \mathbf{W_0}\,.$$

Really, $\mathbf{w}^{(\mathbf{p})}(\mathbf{a}) = \mathbf{0}$, $\mathbf{w}^{(\mathbf{p})}(\mathbf{b}) = \int_{\mathbf{b}}^{\mathbf{a}} \mathbf{p}\mathbf{dx} = \mathbf{0}$ and $\mathbf{w}^{(\mathbf{p})\prime} = \mathbf{p} \in \mathbf{L_2}(\mathbf{a}, \mathbf{b})$. Thus,

$$\sup_{\substack{\mathbf{w} \in \mathbf{W_0} \\ \mathbf{w} \neq \mathbf{0}}} \frac{\int_\Omega \mathbf{p}\,\mathbf{w}'\,\mathbf{dx}}{\|\mathbf{w}'\|} \geq \frac{\int_\Omega \mathbf{p}\,\mathbf{w}^{(\mathbf{p})\prime}\,\mathbf{dx}}{\left\|\mathbf{w}^{(\mathbf{p})\prime}\right\|} = \frac{\int_\Omega \mathbf{p}^2\,\mathbf{dx}}{\|\mathbf{p}\|} = \|\mathbf{p}\|$$

Thus, $\mathbb{C}_\Omega \geq 1$ and we arrive at the required result.

These estimates give a certain presentation on the value of $\mathbb{C}_\Omega$.
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 $\mathbb{C}_\Omega$ by the following arguments.
Let $\mathbf{w_p} \in \mathbf{W_0}$ be a function such that

$$\Delta \mathbf{w_p} \;=\; \nabla \mathbf{p}, \quad \mathbf{w_p} = \mathbf{0} \;\;\textbf{on}\;\; \partial\boldsymbol{\Omega}\,.$$

Then,

$$-\int_{\boldsymbol{\Omega}} \nabla \mathbf{w_p} : \nabla \mathbf{v}\, \mathbf{dx} = \int_{\boldsymbol{\Omega}} \nabla \mathbf{p} \cdot \mathbf{v}\, \mathbf{dx} \quad \forall \mathbf{v} \in \mathbf{W_0}$$

and, thus, we have

$$\int_{\boldsymbol{\Omega}} |\nabla \mathbf{w_p}|^2\, \mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{p}\, \mathbf{div} \mathbf{w_p}\, \mathbf{dx} \quad \forall \mathbf{v} \in \mathbf{W_0}\,.$$

Therefore,

$$\mathbb{C}_{\boldsymbol{\Omega}} := \inf_{\substack{\mathbf{p} \in \overset{\circ}{\mathbf{L}}_2 \\ \mathbf{p} \neq \mathbf{0}}} \sup_{\substack{\mathbf{w} \in \mathbf{W}_0 \\ \mathbf{w} \neq \mathbf{0}}} \frac{\int_{\Omega} \mathbf{p} \operatorname{div} \mathbf{w} \, dx}{\|\mathbf{p}\| \, \|\nabla w\|} \geq \inf_{\substack{\mathbf{p} \in \overset{\circ}{\mathbf{L}}_2 \\ \mathbf{p} \neq \mathbf{0}}} \frac{\int_{\Omega} \mathbf{p} \operatorname{div} \mathbf{w}_{\mathbf{p}} dx}{\|\mathbf{p}\| \, \|\nabla \mathbf{w}_{\mathbf{p}}\|} =$$

$$= \inf_{\substack{\mathbf{p} \in \overset{\circ}{\mathbf{L}}_2 \\ \mathbf{p} \neq \mathbf{0}}} \frac{\|\nabla \mathbf{w}_{\mathbf{p}}\|}{\|\mathbf{p}\|}.$$

Thus, finding $\mathbb{C}_{\Omega}$ requires the minimization of this quotient with respect to all $\mathbf{p} \in \overset{\circ}{\mathbf{L}}_2$, 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. However, the minimization on a subspace of $\overset{\circ}{\mathbf{L}}_2$ may give a presentation on the value of $\mathbb{C}_{\Omega}$.

The value of C_{LBB} is known for several model domains:

- Rectangular domain $(\mathbf{0}, \mathbf{1}) \times (\mathbf{0}, \mathbf{L})$, $\quad \mathbf{L} \geq \mathbf{1}$
  see G. Stoyan, M. Olshanskij, E. Chizhonkov

$$\frac{\sin \frac{\pi}{\mathbf{8}}}{\mathbf{L}} \leq \mathbb{C}_{\mathbf{LBB}} \leq \frac{\pi}{\mathbf{2}\sqrt{\mathbf{3}}\mathbf{L}}$$

- unitary disc with radius 1
  see L. Halpern

$$\mathbb{C}_{\mathbf{LBB}} = \frac{\mathbf{1}}{\sqrt{\mathbf{2}}}$$

Concerning numerical computation of $\mathbb{C}_{LBB}$ see the works of G. Stoyan, M. Olshanskij, E. Chizhonkov

On $\mathbb{C}_\Omega$ for the square domain

Let

$$\Omega = \mathbb{Q} := \{x \in \mathbb{R}^{\mathbf{n}} \mid x_i \in (-\pi, \pi),\ i = 1, 2, \ldots n\}\,.$$

We are interested in the value of the quotient

$$\inf_{\mathbf{p} \in \overset{\circ}{\mathbf{L}}_2} \frac{[\![\, \nabla \mathbf{p} \,]\!]}{\|\mathbf{p}\|_{\mathbb{Q}}}\,.$$

Represent $\mathbf{p}$ as a series with respect to the trial functions

$$\mathbf{p}^{(1)}_{\mathbf{ij}} = \sin \mathbf{ix} \sin \mathbf{jy}, \quad \mathbf{p}^{(2)}_{\mathbf{ij}} = \sin \mathbf{ix} \cos \mathbf{jy},$$

$$\mathbf{p}^{(3)}_{\mathbf{ij}} = \cos \mathbf{ix} \sin \mathbf{jy}, \quad \mathbf{p}^{(4)}_{\mathbf{ij}} = \cos \mathbf{ix} \cos \mathbf{jy}\,,$$

where $i, j = 0, 1, 2, \ldots$ Then

$$\mathbf{p}(\mathbf{x}, \mathbf{y}) = \sum_{\mathbf{i,j=0}}^{\infty} \sum_{\mathbf{s=1}}^{\mathbf{4}} \mathbf{a}^{(\mathbf{s})}_{\mathbf{ij}} \mathbf{p}^{(\mathbf{s})}_{\mathbf{ij}}\,.$$

Here, the first nonzero coefficients are

$$
\begin{aligned}
a_{00}^{(4)} &= \frac{1}{4\pi^2}\int_\Omega p\, dxdy\,, \\
a_{i0}^{(2)} &= \frac{1}{2\pi^2}\int_\Omega p\, \sin ix\, dxdy\,, \\
a_{0j}^{(3)} &= \frac{1}{2\pi^2}\int_\Omega p\, \sin jy\, dxdy\,, \\
a_{i0}^{(4)} &= \frac{1}{2\pi^2}\int_\Omega p\, \cos ix\, dxdy\,, \\
a_{0j}^{(2)} &= \frac{1}{2\pi^2}\int_\Omega p\, \cos jy\, dxdy\,,
\end{aligned}
$$

Other coefficients are as follows:

$$
\begin{aligned}
a_{ij}^{(1)} &= \frac{1}{\pi^2}\int_\Omega p \sin ix \, \sin jy \, dxdy \,, \\
a_{ij}^{(2)} &= \frac{1}{\pi^2}\int_\Omega p \sin ix \, \cos jy \, dxdy \,, \\
a_{ij}^{(3)} &= \frac{1}{\pi^2}\int_\Omega p \cos ix \, \sin jy \, dxdy \,, \\
a_{ij}^{(4)} &= \frac{1}{\pi^2}\int_\Omega p \cos ix \, \cos jy \, dxdy \,.
\end{aligned}
$$

We have

$$
\|\mathbf{p}\|^2_{\mathbb{Q}} = \pi^2 \sum_{\mathbf{i,j=0}}^{\infty} \lambda_{\mathbf{ij}} \left[\left(\mathbf{a}_{\mathbf{ij}}^{(1)}\right)^2 + \left(\mathbf{a}_{\mathbf{ij}}^{(2)}\right)^2 + \left(\mathbf{a}_{\mathbf{ij}}^{(3)}\right)^2 + \left(\mathbf{a}_{\mathbf{ij}}^{(4)}\right)^2 \right] \,,
$$

where $\lambda_{00} = 0$, $\lambda_{01} = 2$, $\lambda_{10} = 2$ and $\lambda_{ij} = 1$ for all $i, j \geq 1$.

Let us take a finite number of elements in the Fourier series for $\mathbf{p}$:

$$\mathbf{p} = \sum_{\mathbf{i,j=0}}^{\mathbf{N}} \sum_{\mathbf{s=1}}^{\mathbf{4}} \mathbf{a_{ij}^{(s)} p_{ij}^{(s)}},$$

where $\mathbf{a_{ij}^{(s)}}$ are the above defined coefficients. Since

$$[\![\nabla \mathbf{p}]\!] = \sup_{\mathbf{v} \in \mathbf{W_0}} \frac{\int_\Omega \mathbf{p}\,\mathbf{div}\mathbf{v}\,\mathbf{dx}}{\|\nabla \mathbf{v}\|_{\mathbb{Q}}}$$

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

$$\mathbf{v} = \mathbf{v}^{\mathbf{M}} = (\mathbf{v}_1^{\mathbf{M}}, \mathbf{v}_2^{\mathbf{M}}),\ \mathbf{v}_1^{\mathbf{M}} = \sum_{\alpha,\beta=1}^{\mathbf{M}} \mathbf{b}_{\alpha\beta}\phi_{\alpha\beta},\ \mathbf{v}_2^{\mathbf{M}} = \sum_{\alpha,\beta=1}^{\mathbf{M}} \mathbf{c}_{\alpha\beta}\phi_{\alpha\beta}\,.$$

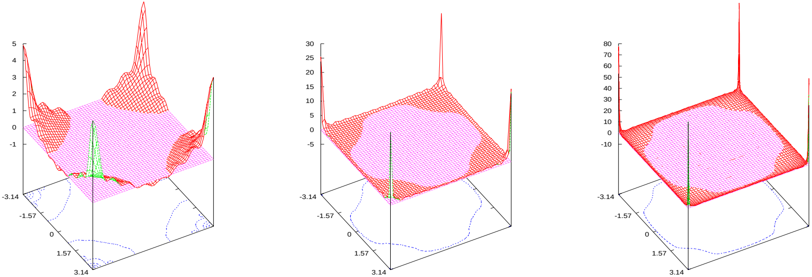
The set of all such functions we denote $\mathbf{W}_0^{\mathbf{M}}$. In this case, we can obtain a lower bound for the required norm. Really, we have

$$[\![\nabla \mathbf{p}]\!]^{(\mathbf{M})} := \sup_{\mathbf{v^M} \in \mathbf{W_0^M}} \frac{\displaystyle\int_\Omega \mathbf{p}\, \mathbf{div} \mathbf{v^M}\, \mathbf{dx}}{\|\nabla \mathbf{v^M}\|\,\mathbb{Q}} \leq [\![\nabla \mathbf{p}]\!] = \sup_{\mathbf{v} \in \mathbf{W_0}} \frac{\displaystyle\int_\Omega \mathbf{p}\, \mathbf{div} \mathbf{v}\, \mathbf{dx}}{\|\nabla \mathbf{v}\|_{\mathbb{Q}}}.$$

Thus, we may hope to estimate the value of the quotient

$$\inf_{\mathbf{p} \in \overset{\circ}{\mathbf{L}}_2} \frac{[\![\nabla \mathbf{p}]\!]}{\|\mathbf{p}\|_{\mathbb{Q}}}.$$

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:

$$\mathbf{div}\boldsymbol{\sigma} = \nabla \mathbf{p} - \mathbf{f} \qquad in \ \ \mathbf{\Omega}, \tag{7.12}$$

$$\mathbf{divu} = \mathbf{0} \qquad in \ \ \mathbf{\Omega}, \tag{7.13}$$

$$\boldsymbol{\sigma} = \boldsymbol{\nu}\nabla \mathbf{u} \qquad in \ \ \mathbf{\Omega}, \tag{7.14}$$

$$\mathbf{u} = \mathbf{0} \qquad on \ \ \partial\mathbf{\Omega}. \tag{7.15}$$

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

$$\mathcal{E} := \overset{\circ}{\mathbf{J}}{}^1_2(\mathbf{\Omega}) \times \mathbf{\Sigma} \times \overset{\circ}{\mathbf{L}}_2 .$$

Problem $\mathcal{P}$. Find $\mathbf{u} \in \overset{\circ}{\mathbf{J}}{}^1_2(\mathbf{\Omega})$ such that

$$\mathbf{J}(\mathbf{u}) \le \mathbf{J}(\mathbf{v}) \quad \text{for all } \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^1_2(\mathbf{\Omega}),$$

where

$$\mathbf{J}(\mathbf{v}) = \int_{\mathbf{\Omega}} \left(\frac{\nu}{2} |\nabla \mathbf{v}|^2 - \mathbf{f} \cdot \mathbf{v} \right) \mathbf{dx}.$$

We denote the exact lower bound of this problem by $\mathit{inf}\,\mathcal{P}$.

Let $\boldsymbol{\Sigma} = \mathbf{L}^2(\boldsymbol{\Omega}, \mathbb{M}^{\mathbf{n}\times\mathbf{n}})$ and $\mathbf{L} : \overset{\circ}{\mathbf{J}}{}^1_2(\boldsymbol{\Omega}) \times \boldsymbol{\Sigma}(\boldsymbol{\Omega}) \to \mathbb{R}$ be the Lagrangian

$$\mathbf{L}(\mathbf{v}, \boldsymbol{\tau}) = \int_{\boldsymbol{\Omega}} \left(\boldsymbol{\tau} : \nabla \mathbf{v} - \frac{1}{2\nu} |\boldsymbol{\tau}|^2 \right) \mathbf{dx} - \int_{\boldsymbol{\Omega}} \mathbf{fv}\, \mathbf{dx}$$

that together with Problem $\mathcal{P}$ generates the dual problem

$$\sup_{\boldsymbol{\tau} \in \boldsymbol{\Sigma}} \inf_{\mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^1_2(\boldsymbol{\Omega})} \mathbf{L}(\mathbf{v}, \boldsymbol{\tau})$$

which is **Problem $\mathcal{P}^*$:** find $\boldsymbol{\sigma} \in \boldsymbol{\Sigma}_f(\boldsymbol{\Omega})$ such that

$$\mathbf{I}^*(\boldsymbol{\sigma}) = \sup_{\boldsymbol{\tau} \in \boldsymbol{\Sigma}_\mathbf{f}(\boldsymbol{\Omega})} \mathbf{I}^*(\boldsymbol{\tau}), \quad \mathbf{I}^*(\boldsymbol{\tau}) = -\frac{1}{2\nu} \int_{\boldsymbol{\Omega}} |\boldsymbol{\tau}|^2 \, \mathbf{dx}$$

where

$$\boldsymbol{\Sigma}_f(\boldsymbol{\Omega}) := \left\{ \boldsymbol{\tau} \in \boldsymbol{\Sigma}(\boldsymbol{\Omega}) \mid \int_{\boldsymbol{\Omega}} \boldsymbol{\tau} : \nabla \mathbf{w}\, \mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{fwdx} \ \text{ for all } \mathbf{w} \in \overset{\circ}{\mathbf{J}}{}^1_2(\boldsymbol{\Omega}) \right\}.$$

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*

$$\mathbf{I}^*(\boldsymbol{\sigma}) = \sup \mathcal{P}^* = \inf \mathcal{P} = \mathbf{I}(\mathbf{u}), \tag{7.16}$$

$$\boldsymbol{\sigma} = \nu \nabla \mathbf{u}\,. \tag{7.17}$$

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* $\boldsymbol{\tau}_f \in \boldsymbol{\Sigma}_f$, *we have*

$$\nu \, \|\nabla(\mathbf{v} - \mathbf{u})\|^2 + \frac{1}{\nu} \|\tau_\mathbf{f} - \sigma\|^2 = 2 \left(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\tau_\mathbf{f}) \right) . \tag{7.18}$$

Proof of Theorem 2

The minimizer $\mathbf{u}$ of problem $\mathcal{P}$ satisfies the relation (**??**). Therefore, we obtain

$$\begin{aligned}
\mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) &= \int_\Omega \left(\frac{\nu}{2}|\nabla\mathbf{v}|^2 - \frac{\nu}{2}|\nabla\mathbf{u}|^2 - \mathbf{f}\cdot(\mathbf{v}-\mathbf{u})\right)\mathbf{dx} = \\
&= \int_\Omega \left(\frac{\nu}{2}|\nabla(\mathbf{v}-\mathbf{u})|^2 + \nu\nabla\mathbf{u} : \nabla(\mathbf{v}-\mathbf{u}) - \mathbf{f}\cdot(\mathbf{v}-\mathbf{u})\right) \\
&= \frac{\nu}{2}\int_\Omega |\nabla(\mathbf{v}-\mathbf{u})|^2\,\mathbf{dx} \qquad \text{for all } \mathbf{v} \in \overset{\circ}{\mathbf{J}}{}^1_2(\Omega).
\end{aligned}$$

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

$$\int_{\Omega} \boldsymbol{\sigma} : (\boldsymbol{\tau}_{\mathbf{f}} - \boldsymbol{\sigma})\, \mathbf{dx} \;=\; \mathbf{0} \qquad \text{for all } \boldsymbol{\tau}_{\mathbf{f}} \in \boldsymbol{\Sigma}_{\mathbf{f}}(\boldsymbol{\Omega})\,.$$

By virtue of this relation, we find that

$$\sup \mathcal{P}^* \;-\; \mathbf{I}^*(\boldsymbol{\tau}_{\mathbf{f}}) \;=\; \mathbf{I}^*(\boldsymbol{\sigma}) - \mathbf{I}^*(\boldsymbol{\tau}_{\mathbf{f}}) \;=\; \frac{\mathbf{1}}{\mathbf{2}\boldsymbol{\nu}}\, \|\boldsymbol{\tau}_{\mathbf{f}} - \boldsymbol{\sigma}_{\mathbf{f}}\|^{\mathbf{2}} \quad \boldsymbol{\tau}_{\mathbf{f}} \in \boldsymbol{\Sigma}_{\mathbf{f}}(\boldsymbol{\Omega}).$$

Since $\inf \mathcal{P} = \sup \mathcal{P}^*$ we sum the two equalities and obtain

$$\boldsymbol{\nu}\, \|\nabla(\mathbf{v} - \mathbf{u})\|^{\mathbf{2}} + \frac{\mathbf{1}}{\boldsymbol{\nu}}\, \|\boldsymbol{\tau}_{\mathbf{f}} - \boldsymbol{\sigma}\|^{\mathbf{2}} = \mathbf{2}\,(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\boldsymbol{\tau}_{\mathbf{f}}))\,.$$

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

$$(\boldsymbol{\mathcal{A}\Lambda}\mathbf{u}, \boldsymbol{\Lambda}\mathbf{w}) + \langle \boldsymbol{\ell}, \mathbf{w} \rangle = \mathbf{0} \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

In this case $\mathbf{V_0} = \overset{\circ}{\mathbf{J}}{}_{\mathbf{2}}^{\mathbf{1}}(\boldsymbol{\Omega})$, $\mathbf{V}$ is a subspace of $\mathbf{H^1}$ containing solenoidal fields, $\boldsymbol{\Lambda} = \nabla$ (tensor–gradient), $\mathbf{U} = \boldsymbol{\Sigma}$, $\boldsymbol{\mathcal{A}}\mathbf{y} = \boldsymbol{\nu}\mathbf{y}$, and

$$\langle \boldsymbol{\ell}, \mathbf{w} \rangle = -\int_{\boldsymbol{\Omega}} \mathbf{f}\mathbf{w}\, \mathbf{dx}$$

Thus, we can apply the estimate

$$\frac{1}{2} \,|\!|\!| \, \Lambda(v-u) \, |\!|\!|^2 \leq (1+\beta) D(\Lambda v, y) + \frac{1+\beta}{2\beta} [\![\, \ell + \Lambda^* y \,]\!]^2, \quad (7.19)$$

where $|\!|\!| \, y \, |\!|\!|^2 = \int_\Omega \nu |y|^2 dx$ and

$$[\![\, \ell + \Lambda^* y \,]\!] = \sup_{w \in V_0} \frac{\langle \ell + \Lambda^* y, w \rangle}{|\!|\!| \, \Lambda w \, |\!|\!|} = \sup_{w \in \overset{\circ}{J}{}^1_2(\Omega)} \frac{\int_\Omega (\nabla w : y - f w) dx}{|\!|\!| \, \nabla w \, |\!|\!|} =$$

$$\sup_{w \in \overset{\circ}{J}{}^1_2(\Omega)} \frac{\int_\Omega (\nabla w : y - f w - q \operatorname{div} w) dx}{|\!|\!| \, \nabla w \, |\!|\!|} \leq$$

$$\leq \sup_{w \in \overset{\circ}{H}{}^1(\Omega)} \frac{\int_\Omega (\nabla w : y - f w - q \operatorname{div} w) dx}{|\!|\!| \, \nabla w \, |\!|\!|} \qquad \forall q \in L^2(\Omega).$$

If

$$\mathbf{y} \in \mathbf{\Sigma}_{\rm div}(\mathbf{\Omega}) := \{\mathbf{y} \in \mathbf{\Sigma} \mid \mathbf{divy} \in \mathbf{L}^2(\mathbf{\Omega}, \mathbb{R}^{\mathbf{n}})\}$$

and $\mathbf{q} \in \mathbf{H}^1$, we have

$$\sup_{\mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^1(\mathbf{\Omega})} \frac{\int_{\mathbf{\Omega}} (\nabla \mathbf{w} : \mathbf{y} - \mathbf{fw} - \mathbf{qdivw})\mathbf{dx}}{|\!|\!| \nabla \mathbf{w} |\!|\!|} = \sup_{\mathbf{w} \in \overset{\circ}{\mathbf{H}}{}^1(\mathbf{\Omega})} \frac{\int_{\mathbf{\Omega}} (\mathbf{f} - \nabla \mathbf{q} + \mathbf{divy}) \cdot \mathbf{wdx}}{|\!|\!| \nabla \mathbf{w} |\!|\!|}$$

Since

$$\|\mathbf{w}\| \leq \mathbf{C}_{\mathbf{\Omega}} \|\nabla \mathbf{w}\| = \mathbf{C}_{\mathbf{\Omega}} \boldsymbol{\nu}^{-1/2} |\!|\!| \nabla \mathbf{w} |\!|\!|,$$

we obtain

$$[\![\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,]\!] \leq \mathbf{C}_{\mathbf{\Omega}} \boldsymbol{\nu}^{-1/2} \|\mathbf{f} - \nabla \mathbf{q} + \mathbf{divy}\|$$

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 (7.19) 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 β we derive the first basic estimate for the Stokes problem:

$$\boxed{\nu\|\nabla(u-v)\| \;\leq\; \|\mathbf{y}-\nu\nabla\mathbf{v}\| + \mathbf{C}_{\mathbf{\Omega}}\|\mathbf{f}-\nabla\mathbf{q}+\mathbf{div}\mathbf{y}\|.} \qquad (7.20)$$

Here $\mathbf{v}$ is **any** conforming approximation of $\mathbf{u}$ and $\mathbf{y}$ is **any** tensor–function in $\mathbf{\Sigma}_{\rm div}(\mathbf{\Omega})$ and $\mathbf{q} \in \mathbf{H}^{\mathbf{1}}$ 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 (7.20) to estimate the first norm in the right–hand side of this inequality.

We obtain

$$\begin{gathered}\boldsymbol{\nu}\|\nabla(\mathbf{u}-\widehat{\mathbf{v}})\| \leq \|\boldsymbol{\nu}\nabla(\widehat{\mathbf{w}}_0+\mathbf{u}_0)-\mathbf{y}\|+\mathbf{C}_\Omega\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\|+\\ +\boldsymbol{\nu}\|\nabla(\widehat{\mathbf{v}}_0-\widehat{\mathbf{w}}_0)\| \leq \|\boldsymbol{\nu}\nabla\widehat{\mathbf{v}}-\mathbf{y}\|+\\ +\mathbf{C}_\Omega\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\|+2\boldsymbol{\nu}\|\nabla(\widehat{\mathbf{v}}_0-\widehat{\mathbf{w}}_0)\|.\end{gathered}$$

Hence, we arrive at the estimate

$$\boxed{\boldsymbol{\nu}\|\nabla(\mathbf{u}-\widehat{\mathbf{v}})\| \leq \|\boldsymbol{\nu}\nabla(\widehat{\mathbf{v}})-\mathbf{y}\|+\mathbf{C}_\Omega\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\|+\frac{2\boldsymbol{\nu}}{\mathbb{C}_\Omega}\|\mathbf{div}\widehat{\mathbf{v}}\|} \tag{7.21}$$

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

$$\boldsymbol{\nu}\,\|\nabla(\mathbf{u}-\widehat{\mathbf{v}})\| \leq \|\boldsymbol{\nu}\nabla(\widehat{\mathbf{v}})-\boldsymbol{\eta}-\mathbf{q}\mathbb{I}\|+\mathbf{C}_{\boldsymbol{\Omega}}\|\mathbf{div}\boldsymbol{\eta}+\mathbf{f}\|+\frac{\mathbf{2}\boldsymbol{\nu}}{\mathbb{C}_{\boldsymbol{\Omega}}}\|\mathbf{div}\widehat{\mathbf{v}}\|. \qquad (7.22)$$

Thus, if the constants $\mathbf{c}_{\boldsymbol{\Omega}}$ and $\mathbb{C}_{\boldsymbol{\Omega}}$ are known (or we know suitable upper bounds for them), then (7.21) and (7.22) 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

$$\boldsymbol{\Sigma}_{\mathbf{k}} \qquad \textbf{and} \qquad \mathbf{Q}_{\mathbf{k}}$$

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. Numerical testing of the estimates has been performed in E. Gorshkova and S. Repin. Error control of the approximate solution to the Stokes equation using a posteriori error estimates of functional type. In *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS 2004, Jyväskylä, 24-28 July, 2004* (electronic).

Estimates for the pressure field

Let $\mathbf{q} \in \overset{\circ}{\mathbf{L}}_{\mathbf{2}}$ be an approximation of the pressure field $\mathbf{p}$. Then $(\mathbf{p}-\mathbf{q}) \in \overset{\circ}{\mathbf{L}}_{\mathbf{2}}$ and the Inf-Sup condition implies the relation

$$\sup_{\mathbf{w} \in \mathbf{V_0},\ \mathbf{w} \neq \mathbf{0}} \frac{\int_{\Omega} (\mathbf{p}-\mathbf{q}) \, \mathbf{div w} \, \mathbf{dx}}{\|\mathbf{p}-\mathbf{q}\| \ \|\nabla \mathbf{w}\|} \geq \mathbb{C}_{\mathbf{\Omega}} \, .$$

Thus, for any small positive ϵ there exists a nonzero function $\mathbf{w}^{\epsilon}_{\mathbf{pq}} \in \mathbf{V_0}$ such that

$$\int_{\mathbf{\Omega}} (\mathbf{p}-\mathbf{q}) \mathrm{div} \mathbf{w}^{\epsilon}_{\mathbf{pq}} \mathbf{dx} \geq (\mathbb{C}_{\mathbf{\Omega}} - \epsilon) \|\mathbf{p}-\mathbf{q}\| \|\nabla \mathbf{w}^{\epsilon}_{\mathbf{pq}}\|.$$

Since

$$\int_\Omega \nu \nabla \mathbf{u} : \nabla \mathbf{w}^\epsilon_{\mathbf{pq}}\, \mathbf{dx} = \int_\Omega \left(\mathbf{f} \cdot \mathbf{w}^\epsilon_{\mathbf{pq}} + \mathbf{p}\, \mathrm{div} \mathbf{w}^\epsilon_{\mathbf{pq}} \right) \mathbf{dx},$$

we have

$$\int_\Omega (\mathbf{p} - \mathbf{q}) \mathbf{div}\, \mathbf{w}^\epsilon_{\mathbf{pq}}\, \mathbf{dx} =$$

$$= \int_\Omega \left\{ \nu \nabla (\mathbf{u} - \widehat{\mathbf{v}}) : \nabla \mathbf{w}^\epsilon_{\mathbf{pq}} + \left(\nu \nabla \widehat{\mathbf{v}} : \nabla \mathbf{w}^\epsilon_{\mathbf{pq}} + \nabla \mathbf{q} \cdot \mathbf{w}^\epsilon_{\mathbf{pq}} - \mathbf{f} \cdot \mathbf{w}^\epsilon_{\mathbf{pq}} \right) \right\} \mathbf{dx}$$

$$= \int_\Omega \nu \nabla (\mathbf{u} - \widehat{\mathbf{v}}) : \nabla \mathbf{w}^\epsilon_{\mathbf{pq}}\, \mathbf{dx} + \int_\Omega \left(\nu \nabla \widehat{\mathbf{v}} - \mathbf{y} : \nabla \mathbf{w}^\epsilon_{\mathbf{pq}} \right) \mathbf{dx}$$

$$+ \int_\Omega \left(\mathbf{y} : \nabla \mathbf{w}^\epsilon_{\mathbf{pq}} + \nabla \mathbf{q} \cdot \mathbf{w}^\epsilon_{\mathbf{pq}} - \mathbf{f} \cdot \mathbf{w}^\epsilon_{\mathbf{pq}} \right) \mathbf{dx},$$

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

$$\|\mathbf{p}-\mathbf{q}\| \leq \frac{1}{(\mathbb{C}_{\boldsymbol{\Omega}}-\epsilon)\|\nabla \mathbf{w}^{\epsilon}_{\mathbf{pq}}\|}$$
$$\times \left[\int_{\boldsymbol{\Omega}} \left(\boldsymbol{\nu}\nabla(\mathbf{u}-\widehat{\mathbf{v}}) : \nabla(\mathbf{w}^{\epsilon}_{\mathbf{pq}}) + (\boldsymbol{\nu}\nabla(\widehat{\mathbf{v}}) - \mathbf{y}) : \nabla(\mathbf{w}^{\epsilon}_{\mathbf{pq}})\right) \mathbf{dx}\right.$$
$$\left. + \int_{\boldsymbol{\Omega}} \left(-\mathbf{w}^{\epsilon}_{\mathbf{pq}} \cdot \mathbf{div}\mathbf{y} + \nabla\mathbf{q} \cdot \mathbf{w}^{\epsilon}_{\mathbf{pq}} - \mathbf{f} \cdot \mathbf{w}^{\epsilon}_{\mathbf{pq}}\right) \mathbf{dx}\right]$$
$$\leq \frac{1}{(\mathbb{C}_{\Omega}-\epsilon)}\left[\boldsymbol{\nu}\|\nabla(\mathbf{u}-\widehat{\mathbf{v}})\| + \|\boldsymbol{\nu}\nabla(\widehat{\mathbf{v}})-\mathbf{y}\| + \mathbf{C}_{\Omega}\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\|\right].$$

The first term in the right–hand side of this inequality is estimated by (7.21).

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:

$$\frac{1}{2}\|\mathbf{p}-\mathbf{q}\| \leq \frac{\nu}{\mathbb{C}_{\Omega}{}^{2}}\|\mathbf{div}\widehat{\mathbf{v}}\| + \tag{7.23}$$

$$+\frac{1}{\mathbb{C}_{\Omega}}\|\nu\nabla(\widehat{\mathbf{v}})-\mathbf{y}\| + \frac{\mathbf{C}_{\Omega}}{\mathbb{C}_{\Omega}}\|\mathbf{div}\mathbf{y}+\mathbf{f}-\nabla\mathbf{q}\|.$$

It is easy to see that the right–hand side of (7.23) consists of the same terms as the right–hand side of (7.21) and vanishes if and only if, $\widehat{\mathbf{v}}=\mathbf{u}$, $\mathbf{y}=\sigma$ and $\mathbf{p}=\mathbf{q}$. However, in this case, the dependence of the penalty multipliers from the constant $\mathbb{C}_{\Omega}$ is stronger.

Problems with condition div$u = \phi$.

In many cases, divergence–free condition is replaced by

$$\mathbf{div u} = \phi \quad \textbf{in } \mathbf{\Omega},$$

where ϕ is a given function in $\overset{\circ}{\mathbf{L}}_\mathbf{2}$. For such functions, we have the problem: find $\mathbf{u}$ that is equal to $\mathbf{u_0}$ on $\partial\Omega$ and

$$-\mathbf{div}\boldsymbol{\sigma} + \nabla \mathbf{p} = \mathbf{f} \quad \textbf{in } \mathbf{\Omega},$$
$$\boldsymbol{\sigma} = \nu \nabla \mathbf{u} \quad \textbf{in } \mathbf{\Omega},$$

Let $\mathbf{u}_\phi \in \mathbf{W_0}$, $\mathbf{div u}_\phi = \phi$. 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 $\bar{\mathbf{u}} \in \overset{\circ}{\mathbf{J}}{}^{\mathbf{1}}_{\mathbf{2}}(\mathbf{\Omega}) + \bar{\mathbf{u}}_\mathbf{0}$ such that

$$-\mathbf{div}\bar{\boldsymbol{\sigma}} + \nabla \mathbf{p} = \bar{\mathbf{f}} \quad \textbf{in } \mathbf{\Omega}, \qquad \bar{\mathbf{f}} = \mathbf{f} + \nu \mathbf{div} \nabla(\mathbf{u}_\phi) \in \mathbf{H}^{-\mathbf{1}},$$
$$\bar{\boldsymbol{\sigma}} = \nu \nabla \bar{\mathbf{u}} \quad \textbf{in } \mathbf{\Omega}.$$

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 (7.21) to a "shifted" system and obtain

$$\|\nabla(\mathbf{u} - \mathbf{v})\| = \|\nabla(\bar{\mathbf{u}} - \bar{\mathbf{v}})\| \leq$$

$$\leq \|\boldsymbol{\nu}\nabla\bar{\mathbf{v}} - \mathbf{y}\| + [\![\, \mathbf{divy} + \bar{\mathbf{f}} - \nabla\mathbf{q} \,]\!] + \frac{2\nu}{\mathbf{C_{LBB}}}\|\mathbf{div}\bar{\mathbf{v}}\|.$$

Set here $\mathbf{y} = -\boldsymbol{\nu}\nabla\mathbf{u}_\phi + \boldsymbol{\eta}$, where $\boldsymbol{\eta}$ is a function in Σ.
Then

$$\mathbf{divy} + \bar{\mathbf{f}} = -\nu\mathbf{div}\nabla\mathbf{u}_\phi + \mathbf{div}\eta + \bar{\mathbf{f}} = \mathbf{div}\eta + \mathbf{f}$$

and $\boldsymbol{\nu}\nabla\bar{\mathbf{v}} - \mathbf{y} = \boldsymbol{\nu}\nabla(\mathbf{v} - \mathbf{u}_\phi) - \mathbf{y} = \boldsymbol{\nu}\nabla\mathbf{v} - \boldsymbol{\eta}$. Therefore,

$$\|\nabla(\mathbf{u} - \mathbf{v})\| \leq$$

$$\leq \|\boldsymbol{\nu}\nabla\mathbf{v} - \eta\| + [\![\, \mathbf{div}\eta + \mathbf{f} - \nabla\mathbf{q} \,]\!] + \frac{2\nu}{\mathbf{C_{LBB}}}\|\mathbf{divv} - \phi\|.$$

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 the term that contains the divergence with a large multiplier. Let us consider a model of such a type.

We find $\mathbf{u}_\delta \in \mathbf{V}$ satisfying the integral identity

$$\int_{\boldsymbol{\Omega}} \left(\boldsymbol{\nu} \nabla \mathbf{u}_\delta : \nabla \mathbf{w} + \frac{\mathbf{1}}{\delta} \mathbf{div} \mathbf{u}_\delta \, \mathbf{div} \mathbf{w}\right) \mathbf{dx} = \int_{\boldsymbol{\Omega}} \mathbf{f} \cdot \mathbf{w} \, \mathbf{dx}, \quad \mathbf{w} \in \mathbf{W_0},$$

and the boundary condition $\mathbf{u}_\delta = \mathbf{u_0}$ $\quad \partial\boldsymbol{\Omega}$. It is not difficult to show (see, e.g., R. Temam [?]), that $\mathbf{u}_\delta$ tends to $\mathbf{u}$ (solution of the Stokes problem) in $\mathbf{H^1}$ norm and $\mathbf{p}_\delta = -\frac{\mathbf{1}}{\delta} \mathbf{div} \mathbf{u}_\delta \in \overset{\circ}{\mathbf{L}}_\mathbf{2}$ converges to the respective pressure function $\mathbf{p}$ in $\mathbf{L_2}$ as $\delta \to \mathbf{0}$.

By (7.21) we can easily obtain an estimate of the difference between $\mathbf{u}$ and $\mathbf{u}_\delta$. Let us set in(7.21) $\mathbf{y} = \boldsymbol{\tau}_\delta := \boldsymbol{\nu}\nabla\mathbf{u}_\delta$ and $\mathbf{q} = \mathbf{p}_\delta = -\frac{\mathbf{1}}{\delta}\mathbf{div}\mathbf{u}_\delta$. In this case, $\|\boldsymbol{\nu}\nabla\mathbf{u}_\delta - \boldsymbol{\tau}_\delta\| = \mathbf{0}$ and

$$[\![\, \mathbf{div}\boldsymbol{\tau}_\delta + \mathbf{f} - \nabla\mathbf{p}_\delta \,]\!] =$$

$$= \sup_{\mathbf{w}\in\mathbf{V_0}} \frac{\int_{\boldsymbol{\Omega}} (-\boldsymbol{\nu}\nabla\mathbf{u}_\delta : \nabla\mathbf{w} + \mathbf{f}\cdot\mathbf{w} + \mathbf{p}_\delta\mathbf{div}\mathbf{w})\, \mathbf{dx}}{\|\nabla\mathbf{w}\|} = \mathbf{0}.$$

Thus, we conclude that

$$\frac{1}{2}\|\nabla(\mathbf{u} - \mathbf{u}_\delta)\| \leq \frac{\mathbf{1}}{\mathbf{C_{LBB}}}\|\mathbf{div}\mathbf{u}_\delta\|,$$

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.

In S. Repin. Estimates of deviations from exact solutions for some boundary–value problems with incompressibility condition. *Algebra and Analiz (St.-Petersburg Math. J)*, 16(2004), 5

functional a posteriori estimates for the Stokes and some other problems were derived by *nonvariational* techniques.
In particular, in this paper readers can find such estimates for
Convection–diffusion equation

$$-\mathbf{div} A\nabla \mathbf{u} + \mathbf{a}\cdot\nabla\mathbf{u} = \mathbf{f}$$

and **Oseen problem**

$$\begin{aligned} -\nu\boldsymbol{\Delta}\mathbf{u} + \mathbf{div}(\mathbf{a}\otimes\mathbf{u}) &= \mathbf{f} - \nabla\mathbf{p} && \textit{in } \Omega,\\ \mathbf{div}\mathbf{u} &= \mathbf{0} && \textit{in } \Omega,\\ \mathbf{u} &= \mathbf{0} && \textit{on } \partial\Omega. \end{aligned}$$

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

$$\|\!| \boldsymbol{\Lambda}(\mathbf{u} - \widehat{\mathbf{v}}) \|\!| \leq$$
$$\leq 2\sqrt{\nu_2}\alpha\|\mathbf{B}\widehat{\mathbf{v}}\| + \|\!| \mathcal{A}\boldsymbol{\Lambda}\widehat{\mathbf{v}} - \mathbf{y} \|\!|_* + \frac{1}{\sqrt{\nu_1}} [\![\mathbf{f} + \boldsymbol{\Lambda}^*\mathbf{y} - \mathbf{B}^*\mathbf{q}]\!] .$$

where $\|\!| \mathbf{y} \|\!| := (\mathcal{A}\mathbf{y}, \mathbf{y})^{1/2}$, $\|\!| \mathbf{y} \|\!|_* := (\mathcal{A}^{-1}\mathbf{y}, \mathbf{y})^{1/2}$ We see that the terms of the estimate present errors in the basic relations

$$\begin{cases} \langle \boldsymbol{\Lambda}^*\sigma + \mathbf{f} - \mathbf{B}^*\mathbf{p}, \mathbf{w} \rangle = 0 \qquad \forall \mathbf{w} \in \mathbf{V_0}, \\ \sigma = \mathcal{A}\boldsymbol{\Lambda}\mathbf{u}, \\ \mathbf{B}\mathbf{v} = \mathbf{0}. \end{cases}$$

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 $\boldsymbol{\nu}_1 = \boldsymbol{\nu}_2 = \boldsymbol{\nu}$,

$$\| \boldsymbol{\mathcal{A}} \mathbf{\Lambda} \widehat{\mathbf{v}} - \mathbf{y} \|_* = \frac{\mathbf{1}}{\sqrt{\boldsymbol{\nu}}} \| \boldsymbol{\nu} \nabla \mathbf{v} - \mathbf{y} \|.$$

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 (7.21).

Literature comments.

A significant part of the difficulties arising in the process of solving such problems is related to the *incompressibility condition*. 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 (see, e.g., A. Chorin [12], E. W. and J. G. Liu [14], V. Girault, P. A. Raviart [16], G. Heywood and R. Rannacher [17], R. Rannacher [25,26], J. Shen [32], R. Temam [33]). 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 [9], F. Brezzi and M. Fortin [10]).

A posteriori error estimates for approximations of the Stokes problem constructed by various types of finite element methods were obtained in numerous papers (mainly in the framework of certain modifications of the residual method (see, e.g., M. Aintworth and T. Oden [1], R. Bank and B. Welfert [5], E. Dari, R. Duran and C. Padra [13], C. Carstensen and S. Funken [11], G. Heywood and R. Rannacher [18], C. Johnson and R. Rannacher [19], C. Johnson, R. Rannacher and M. Boman [20],

J. Oden, W. Wu and M. Aintworth [24], R. Verfürth [35,36]). In these estimates, the right–hand side is given by the sum of local quantities $\eta_{\mathbf{k}}$ that include additional terms that take into account violations of the incompressibility condition.

Functional type a posteriori error estimates for the Stokes problem were firstly derived by the variational techniques in [27]. Later, this method was applied to some other viscous flow problems with non-quadratic dissipative potentials [28]. In [30], another nonvariational techniques was used. It was shown that for the Stokes problem it leads to the same estimates. In [31] a posteriori estimates in local norms were derived.

1. Ainsworth M., Oden J. T., *A posteriori error estimators for the Stokes and Oseen equations*, SIAM J. Numer. Anal. **34** (1997), no. 1, 228–245.
2. Ainsworth M., Oden J. T., *A posteriori error estimation in finite element analysis*, Wiley, New York, 2000.
3. Arnold D. N., Brezzi F., Fortin M., *A stable finite element for the Stokes equations*, Calcolo **21** (1984), no. 4, 337–344 (1985).
4. Babuš ka I., *The finite element method with Lagrangian multipliers*, Numer. Math. **20** (1973), 179–192.
5. Bank R. E., Welfert B. D., *A posteriori error estimates for the Stokes problem*, SIAM J. Numer. Anal. **28** (1991), no. 3, 591–623.
6. Bermudez A., Duran R., Rodriguez R., *Finite element analysis of compressible and incompressible fluid–solid systems*, Math. Comp. **67** (1998), no. 221, 111–136.
7. Bramble J. H., *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.
8. Brezzi F., *On the existence, uniqueness and approximation of saddle–point problems arising from Lagrangian multipliers*, RAIRO Sér. Rouge Anal. Numér. **8** (1974), no. R-2, 129–151.

9 Brezzi F., Douglas J., *Stabilized mixed methods for the Stokes problem*, Numer. Math. **53** (1988), no. 1–2, 225–235.

10 Brezzi F., Fortin M., *Mixed and hybrid finite element methods*, Springer Ser. Comput. Math., vol. 15, Springer-Verlag, New York, 1991.

11 Carstensen C., Funken S. A., *A posteriori error control in low–order finite element discretizations of incompressible stationary flow problems*, Math. Comp. **70** (2001), no. 236, 1353–1381 (electronic).

12 Chorin A. J., *On the convergence of discrete approximations to the Navier–Stokes equations*, Math. Comp. **23** (1969), 341–353.

13 Dari E., Duran R., Padra C., *Error estimators for nonconforming finite element approximations of the Stokes problem*, Math. Comp. **64** (1995), no. 211, 1017–1033.

14 E W., Liu J. G., *Projection method. I. Convergence and numerical boundary layers*, SIAM J. Numer. Anal. **32** (1995), no. 4, 1017–1057.

15 Gorshkova E. and Repin S. Error control of the approximate solution to the Stokes equation using a posteriori error estimates of functional type. In *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS 2004, P.Neittaanmäki, T. Rossi, K. Majava*

and O. Pironeau (eds.), O. Nevanlinna and R. Rannacher (assoc. eds.), Jyväskylä, 24-28 July, 2004 (electronic).

16 Girault V., Raviart P. A., *Finite element approximation of the Navier–Stokes equations*, Lecture Notes in Math., vol. 749, Springer-Verlag, Berlin–New York, 1979.

17 Heywood J. G., Rannacher R., *Finite element approximation of the nonstationary Navier–Stokes problem. I. Regularity of solutions and second-order error estimates for spatial discretization*, SIAM J. Numer. Anal. **19** (1982), 275–311.

18 Heywood J. G., Rannacher R., *Finite-element approximation of the nonstationary Navier–Stokes problem. IV. Error analysis for second-order time discretization*, SIAM J. Numer. Anal. **27** (1990), 353–384.

19 Johnson C., Rannacher R., *On error control in computational fluid dynamics*, Preprint no. 1994-07, Dept. Math. Chalmers Univ. of Technology, Goteborg, 1994.

20 Johnson C., Rannacher R., Boman M., *Numerics and hydrodynamic stability: toward error control in computational fluid dynamics*, SIAM J. Numer. Anal. **32** (1995), 1058–1079.

21 Kobelkov G. M., Olshanskii M., *Effective preconditioning of Uzawa type schemes for a generalized Stokes problem*, Numer. Math. **86** (2000), 443–470.

22 Ladyzhenskaya O.A. Mathematical questions of the dynamics of a viscous incompressible fluid. Second revised and supplemented edition Izdat. "Nauka", Moscow 1970, (in Russian).

23 Ladyzhenskaya O.A. and Solonnikov V.A. Some problems of vector analysis, and generalized formulations of boundary value problems for the Navier-Stokes equation. (Russian) In "Boundary value problems of mathematical physics and related questions in the theory of functions", 9. Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI) 59, (1976), 81–116.

24 Oden J. T., Wu W, Ainsworth M., *An a posteriori error estimate for finite element approximations of the Navier–Stokes equations*, Comput. Methods. Appl. Mech. Engrg. **111** (1994), 185–202.

25 Rannacher R., *Numerical analysis of the Navier–Stokes equations*, Appl. Math. **38** (1993), 361–380.

26 Rannacher R., *Finite element methods for the incompressible Navier–Stokes equations*, Fundamental Directions in Mathematical Fluid

Mechanics (P. Galdi, J. H. Heywood, R Rannacher, eds.), Birkhäuser, Basel, 2000, pp. 191–293.

27 Repin S. *A posteriori estimates for the Stokes problem*, J. Math. Sci. (New York), **109** (2002), 5 1950–1964.

28 Repin S. Estimates of deviations for generalized Newtonian fluids. *Zapiski Nauchn. Semin. V.A. Steklov Mathematical Institute in St.-Petersburg (POMI)*, 288(2002), 178–203.

29 Repin S. Functional type a posteriori error estimates for approximate solutions of problems with incompressibility condition. In *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS 2004, P.Neittaanmäki, T. Rossi, K. Majava and O. Pironeau (eds.), O. Nevanlinna and R. Rannacher (assoc. eds.), Jyväskylä, 24-28 July, 2004* (electronic).

30 Repin S. Estimates of deviations from exact solutions for some boundary–value problems with incompressibility condition. *Algebra and Analiz*, 16(2004), 5, 124–161 (in Russian, English translation in *St.-Petersburg Math. J.*)

31 Repin S. Local a posteriori estimates for the Stokes problem. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 318

(2004), 233–245.

32 Shen J., *On error estimates of the projection methods for the Navier–Stokes equations: second-order schemes*, Math. Comp. **65** (1996), no. 215, 1039–1065.

33 Temam R., *Navier–Stokes equations. Theory and numerical analysis*, Stud. Math. Appl., vol. 2, North–Holland, Amsterdam–New York, 1979.

34 Verfü rth R., *A review of a posteriori error estimation and adaptive mesh–refinement techniques*, Wiley; Teubner, New York, 1996.

35 Verfürth R., *A posteriori error estimators for the Stokes equations*, Numer. Math. **55** (1989), 309–325.

36 Verfürth R., *A posteriori error estimators for the Stokes equations. II. Nonconforming discretizations*, Numer. Math. **60** (1991), 235–249.

Lecture 8.
ESTIMATION OF INDETERMINACY ERRORS.

Errors arising due to data indeterminacy

IN REAL LIFE PROBLEMS ALL THE DATA ARE INDETERMINATE!!!

Example 1.
Diffusion problem: find (temperature) $\mathbf{T}$ such that

$$\begin{aligned} \mathrm{div}\,\mathbf{k}(\mathbf{x})\nabla\mathbf{T}(\mathbf{x}) + \mathbf{f} = \mathbf{0} \qquad & \text{in } \Omega \\ \mathbf{T}(\mathbf{x}) = \mathbf{T_0} \qquad & \text{on } \partial\Omega \end{aligned}$$

In reality, the diffusion coefficient, temperature sources, and even the domain are not exactly known.

Example 2.

Stokes problem: find solenoidal $\mathbf{u}$ such that

$$\begin{aligned} -\,\mathbf{div}\boldsymbol{\sigma} &= \mathbf{f} - \nabla \mathbf{p} && \text{in } \Omega, \\ \nu \nabla \mathbf{u} &= \boldsymbol{\sigma} && \text{in } \Omega, \\ \mathbf{u} &= \mathbf{u_0} && \text{on } \partial\Omega, \end{aligned}$$

In Stokes, Oseen and Navier–Stokes equations the viscosity coefficient is never known exactly. Moreover, it is typically an unknown function (depending both on the spatial and time coordinates) that may depend on the temperature, contamination and other factors.

Uncertain data lead to a quite different analysis

Data uncertainty drastically changes some basic relations in the numerical analysis.

For example, let $\mathbf{v} \in \mathbf{C}^2[\mathbf{a},\mathbf{b}]$ and $\mathbf{x}$, $\mathbf{x}+\mathbf{h} \in [\mathbf{a},\mathbf{b}]$. Since

$$\mathbf{v}(\mathbf{x}+\mathbf{h}) = \mathbf{v}(\mathbf{x}) + \mathbf{v}'(\mathbf{x})\mathbf{h} + \mathbf{v}''(\mathbf{x}+\theta\mathbf{h})\frac{\mathbf{h}^2}{\mathbf{2}}, \quad \theta \in (\mathbf{0},\mathbf{1})$$

we have the standard finite difference quotient approximation of the derivative

$$\mathbf{v}'(\mathbf{x}) \approx \frac{\mathbf{v}(\mathbf{x}+\mathbf{h}) - \mathbf{v}(\mathbf{x})}{\mathbf{h}},$$

whose error is given by the relation

$$|\mathbf{e}| \leq \frac{\boldsymbol{\mu}\mathbf{h}}{\mathbf{2}}, \quad \boldsymbol{\mu} = \max_{\mathbf{x}\in[\mathbf{a},\mathbf{b}]} |\mathbf{v}''(\mathbf{x})|$$

Assume now that $\mathbf{v}(\mathbf{x})$ is defined with a certain indeterminacy, so that its real value is unknown and instead we have a function $\widetilde{\mathbf{v}}(x)$ whose values lie in the interval $[\mathbf{v}(\mathbf{x})-\varepsilon, \mathbf{v}(\mathbf{x})+\varepsilon]$.
If we use these data to approximate the derivative, then we arrive at the following result:

$$\left|\mathbf{v}'(\mathbf{x}) - \frac{\widetilde{\mathbf{v}}(\mathbf{x}+\mathbf{h}) - \widetilde{\mathbf{v}}(\mathbf{x})}{\mathbf{h}}\right| \leq \left|\mathbf{v}'(\mathbf{x}) - \frac{\mathbf{v}(\mathbf{x}+\mathbf{h}) - \mathbf{v}(\mathbf{x})}{\mathbf{h}}\right| + \frac{2\varepsilon}{\mathbf{h}} \leq$$
$$\leq \frac{\mu\mathbf{h}}{\mathbf{2}} + \frac{\mathbf{2}\varepsilon}{\mathbf{h}}.$$

We see that the error does not tend to zero as $\mathbf{h} \to \mathbf{0}$. Moreover,

$$\min_{\mathbf{h}}\left(\frac{\mu\mathbf{h}}{\mathbf{2}} + \frac{\mathbf{2}\varepsilon}{\mathbf{h}}\right)$$

is attained at

$$\bar{\mathbf{h}} = 2\sqrt{\frac{\varepsilon}{\mu}} \qquad (\textbf{the best accuracy}).$$

Therefore, in the case of not fully determinate data the highest accuracy of the numerical differentiation is

$$\boxed{|\mathbf{e}_{\mathbf{min}}| = \mathbf{2}\sqrt{\varepsilon\boldsymbol{\mu}}}$$

For example, if $\varepsilon = 10^{-4}$ and $\boldsymbol{\mu} = 9$, then the highest accuracy is

$$|\mathbf{e}_{\mathbf{min}}| = \mathbf{6} * \mathbf{10}^{-\mathbf{2}} \sim \mathbf{10}^{-\mathbf{1}} \;!!!$$

and it is attained for $\mathbf{h} \approx \mathbf{0.01}$.

Errors in coupled problems

Effects close to those arising as a result of data indeterminacy often appear in the process of numerical simulation of **coupled systems** where certain quantities in a differential problem are defined throughout solutions of some other problems. In such systems a phenomenon of

"error multiplication"

may lead to a dramatic loss of the accuracy. An example below demonstrates such type effects.

"Baby" coupled problem. Find $z(8)$, where z is the solution of the problem

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

where a is a solution of the system of the dimensionality N

$$Ba = f, \quad b_{ij} = \frac{2S_i^2 S_j^2}{\pi} \int\limits_0^{\pi} \left(\sin(i\xi)\sin(j\xi) + \sin(i + j^2)\xi\right) d\xi,$$

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

For $N = 10, 50, 100, 200$ find $z(8)$ analytically and compare with the result obtained by "purely numerical" approach in which sums, integrals and ODE are treated numerically.

Solution.

$$S_i = \sum_{k=0}^{+\infty} \left(\frac{i}{i+1} \right)^k = \sum_{k=0}^{+\infty} q^k = \frac{1}{1-q} = i+1$$

$$Sin(i\xi)Sin(j\xi) = \frac{1}{2}(Cos(i-j)\xi - Cos(i+j)\xi),$$

$$\int\limits_0^\pi Cos((i+j)\xi)d\xi = \frac{1}{i+j} Sin(i+j) \mid_0^\pi = 0,$$

$$\int\limits_0^\pi (Cos(i-j)\xi)d\xi = \pi \; if \; i=j \; and \; = 0 \; otherwise.$$

Therefore,

$$\int\limits_0^\pi \sin(i\xi)\sin(j\xi)d\xi = \frac{\pi}{2} \; if \; i=j \; and \; = 0 \; otherwise.$$

Thus, **B** is a diagonal matrix with

$$\mathbf{b_{ii} = S_i^4 = (i+1)^4}$$

and

$$\mathbf{Ba = f} \quad \text{is} \quad \boxed{\mathbf{(i+1)^4 a_i = (i+1)^4 i}} \Rightarrow \mathbf{a_i = i}.$$

Solution of the equation we find in the form $\mathbf{z = e^{\lambda x}}$, where $\boldsymbol{\lambda}$ is a root of

$$\boldsymbol{\lambda}^2 - 9\boldsymbol{\lambda} - 10 = 0, \qquad \boldsymbol{\lambda}_1 = -1, \ \boldsymbol{\lambda}_2 = 10.$$

We have

$$\mathbf{z = C_1 e^{-x} + C_2 e^{10x}}, \quad \mathbf{z(0) = C_1 + C_2 = 1}$$
$$\mathbf{z' = -C_1 e^{-x} + 10 C_2 e^{10x}}, \quad \mathbf{z'(0) = -C_1 + 10 C_2 = a_{N-1} - a_N = -1}$$

From here, $\mathbf{C_1 = 1}$ and $\mathbf{C_2 = 0}$, so that

$$\mathbf{z = e^{-x}}, \qquad \mathbf{z(8) \approx 3.3546262 * 10^{-4}}$$

Principal question

Assume we have an approximation $\mathbf{u_h}$ computed on a mesh $\mathcal{T}_h$. The question to be answered is as follows:

WHICH ERROR:
APPROXIMATION or INDETERMINACY
IS BIGGER?

If

$$\boxed{\textbf{Indeterminacy error}} > \boxed{\textbf{Approximation error}}$$

then all further computations and mesh adaptations are senseless !
We need a practical way to explicitly evaluate errors caused by indeterminacy in the problem data

General framework

Consider the problem

$$\mathbf{\Lambda}^* \mathcal{A} \mathbf{\Lambda} \mathbf{u} = \boldsymbol{\ell} \ \text{ in } \boldsymbol{\Omega}, \qquad \mathbf{u} = \mathbf{u_0} \text{ on }, \partial\boldsymbol{\Omega}. \tag{8.1}$$

where the operator $\mathcal{A}$ and the functional $\boldsymbol{\ell}$ are defined with some indeterminacy. It means that

$$\begin{aligned} \mathcal{A} \in & \quad \mathcal{U}_{\mathcal{A}} \subset \mathcal{L}(U, U), \\ \boldsymbol{\ell} \in & \quad \mathcal{U}_{\boldsymbol{\ell}} \subset V_0^*, \end{aligned}$$

where $\mathcal{U}_{\mathcal{A}}$ and $\mathcal{U}_{\boldsymbol{\ell}}$ are certain bounded sets.

All possible solutions of the problem with such a data form the set

$$\boldsymbol{\Upsilon}(\mathcal{U}_{\mathcal{A}},\mathcal{U}_{\boldsymbol{\ell}}):=\left\{\widetilde{\mathbf{u}}\in\mathbf{V_0}+\mathbf{u_0}\,\middle|\,\tilde{\mathbf{u}}\,\text{satisfies}\,(8.1)\,\text{for some}\,\mathcal{A}\in\mathcal{U}_{\mathcal{A}}\,\boldsymbol{\ell}\in\mathcal{U}_{\boldsymbol{\ell}}\right\}.$$

Let $\mathbf{v}\in\mathbf{V_0}+\mathbf{u_0}$ be an approximation of an unknown exact solution. Since the data are indeterminate, the error estimation problem comes in two different forms.

The first problem is to find the quantity

$$\mathbf{e}^2_{\min}(\mathbf{v},\boldsymbol{\Upsilon})=\frac{1}{2}\inf_{\tilde{\mathbf{u}}\in\boldsymbol{\Upsilon}}\|\boldsymbol{\Lambda}(\mathbf{v}-\widetilde{\mathbf{u}})\|^2\,,\tag{8.2}$$

The quantity $\mathbf{e}_{\min}$ measures the distance between $\mathbf{v}$ and the set $\boldsymbol{\Upsilon}$. It is equal to zero if $\mathbf{v}$ satisfies (8.1) for some pair $(\mathcal{A},\boldsymbol{\ell})\in\mathcal{U}_{\mathcal{A}}\times\mathcal{U}_{\boldsymbol{\ell}}$. This quantity provides the lowest possible bound of the true error or **the error in the best-case situation**.

Another task is to find the quantity

$$\mathbf{e}^2_{\max}(\mathbf{v}, \boldsymbol{\Upsilon}) = \frac{1}{2} \sup_{\tilde{\mathbf{u}} \in \boldsymbol{\Upsilon}} \|\boldsymbol{\Lambda}(\mathbf{v} - \widetilde{\mathbf{u}})\|^2 , \tag{8.3}$$

which shows the highest possible error. It takes into account computational errors and errors caused by indeterminacy and shows **the error in the worst-case situation** when the exact solution is an element of $\boldsymbol{\Upsilon}$ that is most distant of $\mathbf{v}$.

This quantity is always positive and its value gives an idea of the accuracy limit dictated by the effect of indeterminacy in the data.

Thus,

$$\mathbf{e}_{\min}(\mathbf{v}, \boldsymbol{\Upsilon}) \leq \mathbf{e}(\mathbf{v}) \leq \mathbf{e}_{\max}(\mathbf{v}, \boldsymbol{\Upsilon}), \tag{8.4}$$

where the actual error $\mathbf{e}(\mathbf{v})$ is **principally unknown** and we may only hope to find its bounds. In general, the exact values of $\mathbf{e}_{\min}$ and $\mathbf{e}_{\max}$ could hardly be found. However, using functional type a posteriori estimates, one can find their computable bounds.
Indeed, the majorant $\mathbf{M}_{\oplus}$ and the minorant $\mathbf{M}_{\ominus}$ explicitly depend on $\mathcal{A}$ and ℓ, which opens a way for computing errors caused by the indeterminacy in values of the problem data. Below we show how such an account can be performed.
Assume that the set $\boldsymbol{\Upsilon}$ is known. Our aim is to find **practically computable** numbers $\mathbf{e}_{\ominus}(\mathbf{v}, \boldsymbol{\Upsilon})$ and $\mathbf{e}_{\oplus}(\mathbf{v}, \boldsymbol{\Upsilon})$ such that for any $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$ the following relations hold:

$$\mathbf{e}_{\ominus}(\mathbf{v}, \boldsymbol{\Upsilon}) \leq \mathbf{e}_{\min}(\mathbf{v}, \boldsymbol{\Upsilon}) \ \leq \ \mathbf{e}_{\max}(\mathbf{v}, \boldsymbol{\Upsilon}) \leq \mathbf{e}_{\oplus}(\mathbf{v}, \boldsymbol{\Upsilon}). \tag{8.5}$$

Particular problem

Consider the generalized diffusion problem

$$\mathbf{div}\mathbf{A}\nabla\mathbf{u} + \mathbf{f} = \mathbf{0},$$

In this case,

$$\mathbf{V} = \mathbf{H}^1(\mathbf{\Omega}), \quad \mathbf{Y} \;=\; \mathbf{L}^2\,(\mathbf{\Omega}, \mathbb{R}^{\mathbf{n}}), \quad \mathbf{V_0} := \overset{\circ}{\mathbf{H}}{}^1(\mathbf{\Omega}),$$
$$\mathbf{V}_0^* = \mathbf{H}^{-1}(\mathbf{\Omega}), \quad \mathbf{\Lambda v} = \nabla\mathbf{v}, \quad \mathbf{\Lambda}^*\mathbf{y}^* = -\mathrm{div}\mathbf{y}^*,$$

and $\mathcal{A}$ is a mapping given by the relation $\mathbf{y}^*(\mathbf{x}) \;\rightarrow\; \mathbf{A}(\mathbf{x})\mathbf{y}^*(\mathbf{x})$, where $\mathbf{A}(\mathbf{x})$ is a symmetric positive definite matrix.

Assume the the coefficients of the differential equation are defined by some "mean" elements

$$\mathbf{A_0} \in \mathbf{L}^\infty(\mathbf{\Omega}; \mathbb{M}_\mathbf{s}^{\mathbf{n}\times\mathbf{n}}) \quad \text{and} \quad \mathbf{f_0} \in \mathbf{L}^2(\mathbf{\Omega})$$

and certain (bounded) variations around these values.

$$\mathcal{U}_\mathbf{A} := \left\{\mathbf{A} \in \mathbf{L}^\infty(\mathbf{\Omega}; \mathbb{M}_\mathbf{s}^{\mathbf{n}\times\mathbf{n}}) \mid \mathbf{A} = \mathbf{A_0} + \varepsilon\mathbf{E},\ \mathbf{E} \in \mathcal{E}\right\},$$

$$\mathcal{U}_\mathbf{f} := \left\{\mathbf{f} \in \mathbf{L}^2(\mathbf{\Omega}) \mid \mathbf{f} = \mathbf{f_0} + \delta\varphi,\ \varphi \in \mathcal{F}\right\},$$

where

$$\mathcal{E} := \left\{\mathbf{E} \in \mathbf{L}^\infty(\mathbf{\Omega}; \mathbb{M}_\mathbf{s}^{\mathbf{n}\times\mathbf{n}}) \mid \|\,|\mathbf{E}|\,\|_{\infty,\mathbf{\Omega}} \leq \mathbf{1}\right\},$$

$$\mathcal{F} := \left\{\varphi \in \mathbf{L}^2(\mathbf{\Omega}) \mid \|\varphi\|_{\mathbf{2},\mathbf{\Omega}} \leq \mathbf{1}\right\}.$$

We will define the influence of the above indeterminacy errors. Our analysis follows the lines of

S. Repin. A posteriori error estimates taking into account indeterminacy of the problem data. *Russian J. Numer. Anal.*

Let, ε and δ be small parameters characterizing the range of indeterminacy and

$$\mathcal{U}_{\mathbf{A}} := \left\{\mathbf{A} \in \mathbf{L}^{\infty}(\mathbf{\Omega}; \mathbb{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}}) \mid \mathbf{A} = \mathbf{A_0} + \varepsilon\mathbf{E},\ \mathbf{E} \in \mathcal{E}\right\},$$
$$\mathcal{U}_{\mathbf{f}} := \left\{\mathbf{f} \in \mathbf{L}^{\mathbf{2}}(\mathbf{\Omega}) \mid \mathbf{f} = \mathbf{f_0} + \delta\varphi,\ \varphi \in \mathcal{F}\right\},$$

where

$$\mathcal{E} := \left\{\mathbf{E} \in \mathbf{L}^{\infty}(\mathbf{\Omega}; \mathbb{M}_{\mathbf{s}}^{\mathbf{n}\times\mathbf{n}}) \,\Big|\, \|\,|\mathbf{E}|\,\|_{\infty,\mathbf{\Omega}} \leq \mathbf{1}\right\},$$
$$\mathcal{F} := \left\{\varphi \in \mathbf{L}^{\mathbf{2}}(\mathbf{\Omega}) \,\Big|\, \|\varphi\|_{\mathbf{2},\mathbf{\Omega}} \leq \mathbf{1}\right\}.$$

We assume that the parameter ε is small enough, so that the problems remain uniformly elliptic for all possible data, so that the relation

$$\mathbf{c_1}|\xi|^{\mathbf{2}} \leq \mathbf{A_0}\xi\cdot\xi \leq \mathbf{c_2}|\xi|^{\mathbf{2}}, \qquad \forall\xi \in \mathbb{R}^{\mathbf{n}},$$

implies a similar double inequality for all $\mathbf{A} \subset \mathcal{U}_{\mathbf{A}}$.

Since

$$|\mathbf{E}\xi \cdot \xi| = |\mathbf{E} : (\xi \otimes \xi)| \leq |\mathbf{E}|\, |\xi|^2,$$

we find that

$$\mathbf{A}\xi \cdot \xi \geq \mathbf{A_0}\xi \cdot \xi - \varepsilon|\mathbf{E}|\, |\xi|^2 \geq (\mathbf{c_1} - \varepsilon)|\xi|^2, \tag{8.6}$$

$$\mathbf{A}\xi \cdot \xi \leq \mathbf{A_0}\xi \cdot \xi + \varepsilon|\mathbf{E}|\, |\xi|^2 \leq (\mathbf{c_2} + \varepsilon)|\xi|^2. \tag{8.7}$$

Therefore, we must assume that possible "disturbances" are sufficiently small, namely

$$\varepsilon < \mathbf{c_1}.$$

For the inverse matrix, we have

$$\mathbf{c_2}^{-1}|\xi|^2 \leq \mathbf{A_0}^{-1}\xi \cdot \xi \leq \mathbf{c_1}^{-1}|\xi|^2, \tag{8.8}$$

$$(\mathbf{c_2} + \varepsilon)^{-1}|\xi|^2 \leq \mathbf{A}^{-1}\xi \cdot \xi \leq (\mathbf{c_1} - \varepsilon)^{-1}|\xi|^2, \tag{8.9}$$

where $\mathbf{A} \in \mathcal{U}_{\mathcal{A}}$.

Indeterminacy is explicitly accounted by the Majorant

A principle possibility to involve indeterminacy data into the consideration is based on that **external data are explicitly presented in the Majorant.**

Indeed, we have

$$\int_\Omega \mathbf{A}\nabla(\widetilde{\mathbf{u}}-\mathbf{v})\cdot\nabla(\widetilde{\mathbf{u}}-\mathbf{v})^2 \leq$$

$$(1+\beta)\int_\Omega\left(\mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{v}+\mathbf{A}^{-1}\mathbf{y}\cdot\mathbf{y}-2\nabla\mathbf{v}\cdot\mathbf{y}\right)\mathbf{dx}+$$

$$+\frac{(1+\beta)\mathbf{C}_\Omega^2}{\beta}\left\|\mathbf{div}\mathbf{y}+\mathbf{f}\right\|^2.$$

We do not know $\mathbf{A}$, $\mathbf{f}$ (and also $\mathbf{u}$) exactly. But we can try to express all terms in this estimate via ε, δ, $\mathbf{A_0}$, and $\mathbf{f_0}$.

The left–hand side of the estimate is easy to estimate from below. Indeed,

$$\| \nabla(\mathbf{v} - \widetilde{\mathbf{u}}) \|^2 = \int_\Omega (\mathbf{A_0} + \varepsilon \mathbf{E}) \nabla(\mathbf{v} - \mathbf{u}) \cdot \nabla(\mathbf{v} - \mathbf{u})\, \mathbf{dx} \geq$$
$$\geq (\mathbf{c_1} - \varepsilon) \|\nabla(\mathbf{v} - \mathbf{u})\|^2 .$$

Then, we find that

$$(\mathbf{c_1} - \varepsilon) \|\nabla(\widetilde{\mathbf{u}} - \mathbf{v})\|^2 \leq$$
$$\leq \sup_{\mathbf{A} \in \mathcal{A}, \mathbf{f} \in \mathcal{F}} \inf_{\mathbf{y}, \beta} \Big\{ (1+\beta) \int_\Omega \Big(\mathbf{A} \nabla \mathbf{v} \cdot \nabla \mathbf{v} + \mathbf{A}^{-1} \mathbf{y} \cdot \mathbf{y} - 2 \nabla \mathbf{v} \cdot \mathbf{y} \Big)\, \mathbf{dx} +$$
$$+ \frac{(1+\beta) \mathbf{C}_\Omega^2}{\beta} \|\mathbf{div} \mathbf{y} + \mathbf{f}\|^2 \Big\}.$$

In this estimate an approximate solution v contains both APPROXIMATION and INDETERMINACY Errors !

Basic idea

Since $\sup\inf \leq \inf\sup$, we can change the order and obtain

$$(\mathbf{c_1} - \varepsilon)\,\|\nabla(\mathbf{u}-\mathbf{v})\|^2 \leq$$

$$\leq \inf_{\mathbf{y},\boldsymbol{\beta}} \sup_{\mathbf{A}\in\mathcal{A},\mathbf{f}\in\mathcal{F}} \left\{(\mathbf{1}+\beta)\int_{\Omega} \left(\mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{v} + \mathbf{A}^{-1}\mathbf{y}\cdot\mathbf{y} - 2\nabla\mathbf{v}\cdot\mathbf{y}\right)\mathbf{dx} + \right.$$

$$\left. + \frac{(\mathbf{1}+\beta)\mathbf{C}^2(\Omega,\mathbf{A})}{\beta}\,\|\mathbf{div y} + \mathbf{f}\|^2\right\}.$$

Now, our aim is to find an analytical estimate for the supremum that explicitly involves indeterminacy parameters.

Now, the upper bound of the error of an approximate solution $\mathbf{v}$ with respect to the "worst case situation" comes in the form

$$\mathbf{e}^2_{\max}(\mathbf{v}, \boldsymbol{\Upsilon}) \leq$$
$$\leq \frac{\mathbf{1}}{\mathbf{c_1}}\left(\mathbf{1} + \frac{\varepsilon}{\mathbf{c_1} - \varepsilon}\right)\left\{(\mathbf{1}+\beta) \sup_{\mathbf{A}\in\mathcal{U}_\mathbf{A}} \mathbf{D}(\nabla\mathbf{v}, \mathbf{y}) + \right.$$
$$\left. + \frac{(\mathbf{1}+\beta)}{\mathbf{2}\beta} \sup_{\mathbf{A}\in\mathcal{U}_\mathbf{A}} \mathbf{C}^2(\boldsymbol{\Omega}, \mathbf{A}) \sup_{\mathbf{f}\in\mathcal{U}_\mathbf{f}} \|\operatorname{div}\mathbf{y} - \mathbf{f}\|^2\right\}, \quad (8.10)$$

which is valid for any $\mathbf{y} \in \mathbf{Q}^*$ and $\beta > 0$. Let us consider its terms.

To obtain a transparent estimate we need to find upper bounds for the quantities

$$\sup_{\mathbf{A}\in\mathcal{U}_{\mathbf{A}}} \mathbf{D}(\nabla \mathbf{v}, \mathbf{y}),$$

$$\mathbf{C}^2(\mathbf{\Omega}, \mathbf{A}),$$

$$\sup_{f\in\mathcal{U}_f} \|\mathrm{div}\,\mathbf{y} - \mathbf{f}\|^2.$$

First, we analyze the functional

$$\mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}) := \frac{1}{2}(\mathbf{A}\mathbf{\Lambda v}, \mathbf{\Lambda v}) + \frac{1}{2}(\mathbf{A}^{-1}\mathbf{y}, \mathbf{y}) - (\mathbf{\Lambda v}, \mathbf{y})$$

for any $\mathbf{A} \in \mathcal{U}_{\mathcal{A}}$. First, we rewrite the first term

$$\int_\Omega \mathbf{A}\nabla \mathbf{v} \cdot \nabla \mathbf{v}\, \mathbf{dx} = \int_\Omega (\mathbf{A_0}\nabla \mathbf{v} \cdot \nabla \mathbf{v} + \varepsilon \mathbf{E}\nabla \mathbf{v} \cdot \nabla \mathbf{v})\, \mathbf{dx}.$$

Now, our aim is to estimate the most complicated second term. Present the inverse matrix as follows

$$\mathbf{A}^{-1} = (\mathbf{A_0} + \varepsilon \mathbf{E})^{-1} = \left(\mathbf{A_0}(\mathbb{I} + \varepsilon \mathbf{A_0^{-1}E})\right)^{-1} = (\mathbb{I} + \varepsilon \mathbf{B})^{-1}\, \mathbf{A_0^{-1}},$$

where $\mathbf{B} = \mathbf{A_0^{-1}E}$. Note that

$$\varepsilon|\mathbf{B}| = \varepsilon|\mathbf{A_0^{-1}E}| \le \varepsilon|\mathbf{A_0^{-1}}|\,|\mathbf{E}| \le \varepsilon \mathbf{c_1^{-1}} < \mathbf{1},$$

and, therefore, $(\mathbb{I} + \varepsilon \mathbf{B})^{-1}$ can be presented as a convergent matrix series, namely

∞

$$(\mathbb{I} + \varepsilon \mathbf{B})^{-1} = \mathbb{I} + \sum_{j=1}^{\infty} (-1)^j \varepsilon^j \mathbf{B}^j,$$

Hence, we can present the second term as a combination of known matrixes $\mathbf{A_0}$ and powers of ε.

$$\mathbf{A}^{-1}\mathbf{y}\cdot\mathbf{y} = (\mathbb{I}+\varepsilon\mathbf{B})^{-1}\mathbf{A}_0^{-1}\mathbf{y}\cdot\mathbf{y} = \left(\mathbb{I} + \sum_{j=1}^{\infty}(-1)^j\varepsilon^j\mathbf{B}^j\right)\mathbf{A}_0^{-1}\mathbf{y}\cdot\mathbf{y} =$$

$$= \mathbf{A}_0^{-1}\mathbf{y}\cdot\mathbf{y} - \varepsilon\mathbf{B}\mathbf{A}_0^{-1}\mathbf{y}\cdot\mathbf{y} + \sum_{j=2}^{\infty}(-1)^j\varepsilon^j\mathbf{B}^j\mathbf{A}_0^{-1}\mathbf{y}\cdot\mathbf{y}.$$

Since $\mathbf{E} \in \mathcal{E}$, we have

$$\int_{\Omega} \mathbf{B}^{\mathbf{j}} \mathbf{A}_0^{-1} \mathbf{y} \cdot \mathbf{y} \, \mathbf{dx} \leq \int_{\Omega} |\mathbf{A}_0^{-1}|^{\mathbf{j}+1} \, |\mathbf{E}|^{\mathbf{j}} \, |\mathbf{y}|^2 \, \mathbf{dx} \leq \mathbf{c}_1^{-(\mathbf{j}+1)} \|\mathbf{y}\|^2$$

$$\int_{\Omega} \mathbf{A}^{-1} \mathbf{y} \cdot \mathbf{y} \, \mathbf{dx} \leq \int_{\Omega} (\mathbf{A}_0^{-1} \mathbf{y} \cdot \mathbf{y} - \varepsilon \mathbf{B} \mathbf{A}_0^{-1} \mathbf{y} \cdot \mathbf{y}) \, \mathbf{dx} +$$

$$+ \left(\sum_{\mathbf{j}=2}^{\infty} (-1)^{\mathbf{j}} \varepsilon^{\mathbf{j}} \mathbf{c}_1^{-(\mathbf{j}+1)} \right) \|\mathbf{y}\|^2.$$

We find that the first term **D** explicitly depends on ε:

$$\mathbf{D}(\nabla \mathbf{v}, \mathbf{y}) \leq \mathbf{D_0}(\nabla \mathbf{v}, \mathbf{y}) + \frac{\varepsilon}{2} \int_\Omega (\mathbf{E} \nabla \mathbf{v} \cdot \nabla \mathbf{v} - \mathbf{B} \mathbf{A}_0^{-1} \mathbf{y} \cdot \mathbf{y})\, \mathbf{dx} + \\ + \frac{1}{2\mathbf{c_1}} \|\mathbf{y}\|^2 \sum_{\mathbf{j}=2}^{\infty} \left(-\frac{\varepsilon}{\mathbf{c_1}} \right)^{\mathbf{j}},$$

where

$$\mathbf{D_0}(\nabla \mathbf{v}, \mathbf{y}) = \int_\Omega \left(\frac{1}{2} \mathbf{A_0} \nabla \mathbf{v} \cdot \nabla \mathbf{v} + \frac{1}{2} \mathbf{A}_0^{-1} \mathbf{y} \cdot \mathbf{y} - \nabla \mathbf{v} \cdot \mathbf{y} \right) \mathbf{dx}.$$

Since all the matrices are symmetric, we have

$$\mathbf{E}\nabla\mathbf{v}\cdot\nabla\mathbf{v}-\mathbf{A}_0^{-1}\mathbf{E}\mathbf{A}_0^{-1}\mathbf{y}\cdot\mathbf{y}=\mathbf{E}(\nabla\mathbf{v}-\mathbf{A}_0^{-1}\mathbf{y})\cdot(\nabla\mathbf{v}+\mathbf{A}_0^{-1}\mathbf{y}).$$

Now, we obtain

$$\begin{aligned}\mathbf{D}(\nabla\mathbf{v},\mathbf{y})\le\mathbf{D_0}(\nabla\mathbf{v},\mathbf{y})+&\\ +\frac{\varepsilon}{2}\int_\Omega \mathbf{E}(\nabla\mathbf{v}-\mathbf{A}_0^{-1}\mathbf{y})\cdot(\nabla\mathbf{v}+\mathbf{A}_0^{-1}\mathbf{y})\,\mathbf{dx}+&\\ +\left(\frac{\varepsilon}{\mathbf{c_1}}\right)^2\frac{1}{2(\varepsilon+\mathbf{c_1})}\|\mathbf{y}\|^2.&\end{aligned}$$

Note that the the last two terms presents a positive **penalty arose due to indeterminacy**. All the terms in the right–hand side are **directly computable**!

Next,

$$\frac{1}{C^2(\Omega, A)} = \inf_{w \in V_0} \frac{\int_\Omega A\nabla w \cdot \nabla w \, dx}{\|w\|^2},$$

where

$$\int_\Omega A\nabla w \cdot \nabla w \, dx \geq \int_\Omega A_0 \nabla w \cdot \nabla w \, dx - \varepsilon \|\nabla w\|^2.$$

Hence,

$$\frac{1}{C^2(\Omega, A)} \geq (1 - \varepsilon c_1^{-1}) \inf_{w \in V_0} \frac{\int_\Omega A_0 \nabla w \cdot \nabla w \, dx}{\|w\|^2} = \frac{c_1 - \varepsilon}{c_1} \frac{1}{C^2(\Omega, A_0)}$$

and

$$C^2(\Omega, A) \leq \left(1 + \frac{\varepsilon}{c_1 - \varepsilon}\right) C^2(\Omega, A_0).$$

For any $\mathbf{g} \in \mathbf{L}^2(\boldsymbol{\Omega})$

$$\sup_{\varphi \in \mathcal{F}} \int_{\boldsymbol{\Omega}} (\mathbf{g} - \varphi)^2 \mathbf{dx} = \|\mathbf{g}\|^2 + 2\|\mathbf{g}\| + 1.$$

By this relation, we find the value of the term

$$\sup_{\varphi \in \mathcal{F}} \int_{\boldsymbol{\Omega}} |\mathbf{div}\,\mathbf{y} - \mathbf{f_0} - \delta\varphi|^2 \mathbf{dx},$$

which is

$$\|\mathbf{div}\,\mathbf{y} - \mathbf{f_0}\|^2 - 2\delta\|\mathbf{div}\,\mathbf{y} - \mathbf{f_0}\| + \delta^2.$$

Then, we arrive at the final estimate. To represent its right-hand side of this estimate in a more transparent form, we introduce a number of quantities.

$$
\begin{aligned}
\mathbf{M_{00}}(\mathbf{v},\beta,\mathbf{y}) &:= (\mathbf{1}+\beta)\mathbf{D_0}(\nabla\mathbf{v},\mathbf{y}) + \left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\frac{\mathbf{C}^2(\boldsymbol{\Omega},\mathbf{A_0})}{\mathbf{2}}\|\operatorname{div}\mathbf{y}-\mathbf{f_0}\|^2,\\
\mathbf{M_{10}}(\mathbf{v},\beta,\mathbf{y}) &:= \frac{\varepsilon}{\mathbf{2}}\Big(\int_{\boldsymbol{\Omega}}\left|(\nabla\mathbf{v}-\mathbf{A_0^{-1}}\mathbf{y})\cdot(\nabla\mathbf{v}+\mathbf{A_0^{-1}}\mathbf{y})\right|\mathbf{dx}+\\
&\quad+\left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\frac{\mathbf{C}^2(\boldsymbol{\Omega},\mathbf{A_0})}{\mathbf{c_1}-\varepsilon}\int_{\boldsymbol{\Omega}}|\operatorname{div}\mathbf{y}-\mathbf{f_0}|^2\,\mathbf{dx}\Big),\\
\mathbf{M_{01}}(\mathbf{v},\beta,\mathbf{y}) &:= \delta\left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\mathbf{C}^2(\boldsymbol{\Omega},\mathbf{A_0})\|\operatorname{div}\mathbf{y}-\mathbf{f_0}\|,\\
\mathbf{M_{11}}(\mathbf{v},\beta,\mathbf{y}) &:= \varepsilon\delta\left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\frac{\mathbf{C}^2(\boldsymbol{\Omega},\mathbf{A_0})}{\mathbf{c_1}-\varepsilon}\|\operatorname{div}\mathbf{y}-\mathbf{f_0}\|,\\
\mathbf{M_{22}}(\mathbf{v},\beta,\mathbf{y}) &:= (\mathbf{1}+\beta)\left(\frac{\varepsilon}{\mathbf{c_1}}\right)^{\mathbf{2}}\frac{\|\mathbf{y}\|^2}{\mathbf{2}(\varepsilon+\mathbf{c_1})}+\left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\frac{\mathbf{c_1}\mathbf{C}^2(\boldsymbol{\Omega},\mathbf{A_0})}{\mathbf{2}(\mathbf{c_1}-\varepsilon)}\delta^2.
\end{aligned}
$$

We obtain an upper bound of $\mathbf{e}_{\oplus}(\mathbf{v}, \boldsymbol{\Upsilon})$ in the form

$$\mathbf{e}^2_{\oplus}(\mathbf{v}, \boldsymbol{\Upsilon}) = \frac{\mathbf{1}}{\mathbf{c_1} - \varepsilon} \inf_{\mathbf{y} \in \mathbf{Q}^*, \beta > \mathbf{0}} \Big(\mathbf{M_{00}}(\mathbf{v}, \beta, \mathbf{y}) + \\ + \mathbf{M_{01}}(\mathbf{v}, \beta, \mathbf{y}) + \mathbf{M_{10}}(\mathbf{v}, \beta, \mathbf{y}) + \mathbf{M_{11}}(\mathbf{v}, \beta, \mathbf{y}) + \mathbf{M_{22}}(\mathbf{v}, \beta, \mathbf{y}) \Big). \tag{8.11}$$

The term $\mathbf{M_{00}}(\mathbf{v}, \beta, \mathbf{y})$ coincides with the majorant constructed for the "mean" problem (with $\mathbf{A_0}$ and $\mathbf{f_0}$). It represents the **approximation error**. The terms $\mathbf{M_{10}}$, $\mathbf{M_{01}}$, and $\mathbf{M_{11}}$ are given by some combinations of the weighted residual and small parameters ε and δ. In principle, all these terms can be made arbitrarily small by taking $\mathbf{v}$ close enough to the exact solution $\mathbf{u}$ of the problem with $\mathbf{A} = \mathbf{A_0}$ and $\ell = \mathbf{f_0}$ and $\mathbf{y}$ close enough to $\mathbf{A}_0 \nabla \mathbf{u}$.

Inherent error

In contrast, the term $\mathbf{M_{22}}(\mathbf{v}, \beta; \mathbf{y})$ is always positive. This term contains the **inherent** part of the error, which does not depend on the accuracy of numerical approximations. Indeed, in all cases we have

$$\mathbf{M_{22}}(\mathbf{v}, \beta, \mathbf{y}) \geq \mathbf{C}^2(\Omega, \mathbf{A_0}) \frac{\mathbf{c_1}\delta^2}{2(\mathbf{c_1} - \varepsilon)}.$$

This quantity does not depend on the choice of $\mathbf{v}$, β, and $\mathbf{y}$. It gives an idea of the accuracy limit that could be achieved within the framework of the worst-case scenario.

Computable upper bounds

Take $\{\mathbf{Q}^*_\mathbf{k}\} \subset \mathbf{Q}^*$. Then,

$$\mathbf{e}^2_\oplus(\mathbf{v},\boldsymbol{\Upsilon}) \leq \mathbf{e}^2_{\mathbf{k}\oplus}(\mathbf{v},\boldsymbol{\Upsilon}) =$$

$$= \frac{1}{\mathbf{c_1}}\left(\mathbf{1}+\frac{\varepsilon}{\mathbf{c_1}-\varepsilon}\right) \inf_{\mathbf{y}\in\mathbf{Q}^*_\mathbf{k},\beta>\mathbf{0}} \left\{ \sum_{\mathbf{s},\mathbf{t}=\mathbf{0}}^{\mathbf{1}} \mathbf{M_{st}}(\mathbf{v},\beta,\mathbf{y}) + \mathbf{M_{22}}(\mathbf{v},\beta,\mathbf{y}) \right\}.$$

If $\mathbf{Q}^*_\mathbf{k} \subset \mathbf{Q}^*_{\mathbf{k+1}}$, then the sequence $\{\mathbf{e}^2_{\mathbf{k}\oplus}(\mathbf{v},\boldsymbol{\Upsilon})\}$ monotonically decreases but may not tend to zero.

Lower bound of the error

To find a lower bound, we use the relation

$$\frac{1}{2}\int_\Omega \mathbf{A}\nabla(\mathbf{v}-\widetilde{\mathbf{u}})\cdot\nabla(\mathbf{v}-\mathbf{u})\,\mathbf{dx} = \sup_{\mathbf{w}\in\mathbf{V_0}} \mathbf{M}_\ominus(\mathbf{v},\mathbf{w}),$$

where

$$\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}) = -\int_\Omega \left(\frac{1}{2}\mathbf{A}\nabla\mathbf{w}\cdot\nabla\mathbf{w} + \mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{w} + \mathbf{f}\mathbf{w}\right)\mathbf{dx}.$$

Recall that

$$\mathbf{A}\xi\cdot\xi \le (\varepsilon + \mathbf{c_2})\,|\xi|^2\,.$$

By this inequality we can estimate the left–hand side from below, so that for any pair

$$(\mathbf{A},\mathbf{f}) \in \mathcal{U}_\mathbf{A} \times \mathcal{U}_\mathbf{f}$$

that generates the respective solution $\widetilde{\mathbf{u}}$ we have

$$\frac{1}{2}\|\nabla(\mathbf{v}-\tilde{\mathbf{u}})\|^2 \geq \frac{1}{2(\varepsilon+\mathbf{c}_2)}\int_\Omega \mathbf{A}\nabla(\mathbf{v}-\tilde{\mathbf{u}})\cdot\nabla(\mathbf{v}-\tilde{\mathbf{u}})\,\mathbf{dx} = $$
$$= \frac{1}{\varepsilon+\mathbf{c}_2}\sup_{\mathbf{w}\in\mathbf{V}_0}\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}).$$

Therefore,

$$\mathbf{e}_{\min}(\mathbf{v},\boldsymbol{\Upsilon}) = \inf_{\tilde{\mathbf{u}}\in\boldsymbol{\Upsilon}}\frac{1}{2}\|\nabla(\mathbf{v}-\tilde{\mathbf{u}})\|^2 \geq$$
$$\geq \frac{1}{\varepsilon+\mathbf{c}_2}\inf_{(\mathbf{A},\mathbf{f})\in\mathcal{U}_\mathbf{A}\times\mathcal{U}_\mathbf{f}}\sup_{\mathbf{w}\in\mathbf{V}_0}\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}) \geq$$
$$\geq \frac{1}{\varepsilon+\mathbf{c}_2}\sup_{\mathbf{w}\in\mathbf{V}_0}\inf_{(\mathbf{A},\mathbf{f})\in\mathcal{U}_\mathbf{A}\times\mathcal{U}_\mathbf{f}}\mathbf{M}_\ominus(\mathbf{v},\mathbf{w}).$$

We have

$$
\begin{aligned}
&\mathbf{e}_{\min}(\mathbf{v}, \boldsymbol{\Upsilon}) \geq \\
&\geq \frac{1}{\varepsilon + \mathbf{c_2}} \sup_{\mathbf{w} \in \mathbf{V_0}} \left\{ -\int_{\Omega} \left(\frac{1}{2} \mathbf{A_0} \nabla \mathbf{w} \cdot \nabla \mathbf{w} + \mathbf{A_0} \nabla \mathbf{v} \cdot \nabla \mathbf{w} + \mathbf{f_0} \mathbf{w} \right) \mathbf{dx} + \right. \\
&\left. + \inf_{\mathbf{E} \in \mathcal{E}, \varphi \in \mathcal{F}} \left(-\varepsilon \int_{\Omega} \left(\frac{1}{2} \mathbf{E} \nabla \mathbf{w} \cdot \nabla \mathbf{w} + \mathbf{E} \nabla \mathbf{v} \cdot \nabla \mathbf{w} \right) \mathbf{dx} - \delta \int_{\Omega} \mathbf{w} \varphi \, \mathbf{dx} \right) \right\}.
\end{aligned}
$$

By the algebraic inequality

$$\mathbf{Ea} \cdot \mathbf{b} + \mathbf{Ec} \cdot \mathbf{b} = \mathbf{E_{ij}}(\mathbf{a_j b_j} + \mathbf{c_j b_i}) = \mathbf{E_{ij}}(\mathbf{b_i}(\mathbf{a_j} + \mathbf{c_j})) = \mathbf{E} : (\mathbf{b} \otimes (\mathbf{a} + \mathbf{c}))$$

we find that

$$\inf_{\mathbf{E} \in \mathcal{E}} \left\{ -\int_{\Omega} \left(\frac{1}{2} \mathbf{E} \nabla \mathbf{w} \cdot \nabla \mathbf{w} + \mathbf{E} \nabla \mathbf{v} \cdot \nabla \mathbf{w} \right) \mathbf{dx} \right\} = \\ = -\int_{\Omega} \left| \left(\frac{1}{2} \nabla \mathbf{w} + \nabla \mathbf{v} \right) \otimes \nabla \mathbf{w} \right| \mathbf{dx}.$$

It is easy to see that

$$\inf_{\varphi \in \mathcal{F}} \left\{ -\int_{\Omega} \mathbf{w} \varphi \, \mathbf{dx} \right\} = -\|\mathbf{w}\|.$$

Now, we obtain

$$\mathbf{e}^2_{\min}(\mathbf{v},\boldsymbol{\Upsilon}) \geq$$

$$\geq \frac{1}{\varepsilon+\mathbf{c}_2}\sup_{\mathbf{w}\in\mathbf{V}_0}\left\{-\int_\Omega\left(\frac{1}{2}\mathbf{A}_0\nabla\mathbf{w}\cdot\nabla\mathbf{w}+\mathbf{A}_0\nabla\mathbf{v}\cdot\nabla\mathbf{w}+\mathbf{f}_0\mathbf{w}\right)\,\mathbf{dx}-\right.$$

$$-\,\varepsilon\int_\Omega\left|\left(\frac{1}{2}\nabla\mathbf{w}+\nabla\mathbf{v}\right)\otimes\nabla\mathbf{w}\right|\mathbf{dx}-$$

$$\left.-\,\delta\|w\|\right\}.$$

Introduce the quantities

$$\begin{aligned}
\mathbf{m_{00}}(\mathbf{v},\mathbf{w}) = & \ -\int_\Omega \left(\frac{1}{2}\mathbf{A_0}\nabla\mathbf{w}\cdot\nabla\mathbf{w}+\mathbf{A_0}\nabla\mathbf{v}\cdot\nabla\mathbf{w}+\mathbf{f_0}\mathbf{w}\right)\mathbf{dx},\\
\mathbf{m_{10}}(\mathbf{v},\mathbf{w}) = & \ -\varepsilon\int_\Omega \left|\left(\frac{1}{2}\nabla\mathbf{w}\otimes\nabla\mathbf{w}+\nabla\mathbf{v}\otimes\nabla\mathbf{w}\right)\right|\mathbf{dx},\\
\mathbf{m_{01}}(\mathbf{w}) = & \ -\delta\|\mathbf{w}\|.
\end{aligned}$$

Then, we represent the lower bound in the form

$$\mathbf{e}^2_{\ominus}(\mathbf{v},\boldsymbol{\Upsilon}) := \frac{\mathbf{1}}{\varepsilon+\mathbf{c_2}}\sup_{\mathbf{w}\in\mathbf{V_0}}\{\mathbf{m_{00}}(\mathbf{v},\mathbf{w})+\mathbf{m_{01}}(\mathbf{v},\mathbf{w})+\mathbf{m_{10}}(\mathbf{w})\}\geq\mathbf{0}. \tag{8.12}$$

In this estimate, the term $\mathbf{m_{00}}(\mathbf{v},\mathbf{w})$ contains the major part of the approximation error. It vanishes if $\mathbf{v}$ is a solution of the "mean" problem with $\mathbf{A}=\mathbf{A_0}$ and $\ell=f_0$. Two other terms reflect the influence of the small parameters δ and ε.

Computable lower bounds

Take a collection of finite-dimensional subspace $\mathbf{V_{0k}}$ and solve the problems

$$\mathbf{e}^2_{\min}(\mathbf{v},\boldsymbol{\Upsilon}) \geq \mathbf{e}^2_{\mathbf{k}\ominus}(\mathbf{v},\boldsymbol{\Upsilon}) =$$

$$= \frac{\mathbf{1}}{\varepsilon + \mathbf{c_2}} \sup_{\mathbf{w}\in\mathbf{V_{0k}}} \{\mathbf{m_{00}}(\mathbf{v},\mathbf{w}) + \mathbf{m_{01}}(\mathbf{v},\mathbf{w}) + \mathbf{m_{10}}(\mathbf{w})\}.$$

Now $\mathbf{e}^2_{\mathbf{k}\ominus}(\mathbf{v},\boldsymbol{\Upsilon})$ can be used to estimate the efficiency of further computational efforts within the framework of the best-case scenario.

To refine or not to refine? That is the question.

If $\mathbf{e}^2_{\mathbf{k}\ominus}(\mathbf{v}, \boldsymbol{\Upsilon})$ are large, then **approximation errors are significant**. In this case, it is worth computing a new approximation on a finer mesh.

If for a certain $\mathbf{k}$ the quantity $\mathbf{e}^2_{\mathbf{k}\ominus}(\mathbf{v}, \boldsymbol{\Upsilon})$ is very small, then an approximate solution computed is already close to some $\mathbf{u} \in \boldsymbol{\Upsilon}$. Since we do not know exactly the data (and, thus, have no way to select the proper $\mathbf{u}$) **all further computations and mesh refinements are in a sense useless** because they cannot improve our presentation on the true solution.

Lecture 9.
A POSTERIORI ESTIMATES FOR MIXED METHODS

Mixed approximations. A glance from the minimax theory

Consider our basic problem

$$\begin{aligned}
&\mathbf{div A \nabla u + f = 0} \quad \text{in } \boldsymbol{\Omega}\,,\\
&\mathbf{u = u_0} \text{on } \partial_1 \boldsymbol{\Omega}\,,\\
&\mathbf{A \nabla u \cdot n = F} \text{on } \partial_2 \boldsymbol{\Omega}\,,
\end{aligned}$$

$$\mathbf{c_1^2} |\xi|^2 \leq \mathbf{A(x)} \xi \cdot \xi \leq \mathbf{c_2^2} |\xi|^2 \qquad \forall \xi \in \mathbb{R}^{\mathbf{d}}\,, \text{ for a.e. } \mathbf{x} \in \boldsymbol{\Omega}\,,$$

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

$$\begin{aligned}
&\mathbf{V} := \mathbf{H^1}(\boldsymbol{\Omega})\,,\ \mathbf{V_0} := \{\mathbf{v} \in \mathbf{V} \mid \mathbf{v = 0} \text{ on } \partial_1 \boldsymbol{\Omega}\}\,,\ \widehat{\mathbf{V}} := \mathbf{L_2}(\boldsymbol{\Omega})\,,\\
&\mathbf{Q} := \mathbf{L_2}(\boldsymbol{\Omega}; \mathbb{R}^{\mathbf{d}}) \qquad \widehat{\mathbf{Q}} := \mathbf{H}(\boldsymbol{\Omega}; \mathbf{div})\,,\\
&\widehat{\mathbf{Q}}^+ := \{\mathbf{y} \in \widehat{\mathbf{Q}} \mid \mathbf{y} \cdot \mathbf{n}\big|_{\partial_2 \boldsymbol{\Omega}} \in \mathbf{L_2}(\partial_2 \boldsymbol{\Omega})\}\,.
\end{aligned}$$

We recall that $\|\mathbf{q}\|_{\mathbf{div}}$ is the norm in $H(\boldsymbol{\Omega};\mathbf{div})$:

$$\|\mathbf{q}\|_{\mathbf{div}} := (\|\mathbf{q}\|^2 + \|\mathbf{div}\mathbf{q}\|^2)^{1/2} \quad \forall \mathbf{q} \in \mathbf{Q}$$

and

$$\| \mathbf{q} \| := \left(\int\limits_{\boldsymbol{\Omega}} \mathbf{A}\mathbf{q} \cdot \mathbf{q}\, \mathbf{dx} \right)^{1/2}, \quad \mathbf{q} \in \mathbf{Q}$$

$$\| \mathbf{q} \|_* := \left(\int\limits_{\boldsymbol{\Omega}} \mathbf{A}^{-1}\mathbf{q} \cdot \mathbf{q}\, \mathbf{dx} \right)^{1/2}$$

Note that,

$$\bar{\mathbf{c}}_1^2|\xi|^2 \leq \mathbf{A}^{-1}(\mathbf{x})\xi \cdot \xi \leq \bar{\mathbf{c}}_2^2|\xi|^2 \qquad \forall \xi \in \mathbb{R}^{\mathbf{d}}, \text{ for a.e. } \mathbf{x} \in \boldsymbol{\Omega}$$

with $\bar{\mathbf{c}}_1 = \mathbf{1}/\mathbf{c}_2$, $\bar{\mathbf{c}}_2 = \mathbf{1}/\mathbf{c}_1$.

Generalized solution of the problem considered can be viewed as a saddle point of the Lagrangian

$$\mathbf{L}(\mathbf{v},\mathbf{q}) := \int_{\mathbf{\Omega}} \left(\nabla \mathbf{v} \cdot \mathbf{q} - \frac{\mathbf{1}}{\mathbf{2}} \mathbf{A}^{-\mathbf{1}} \mathbf{q} \cdot \mathbf{q} \right) \, \mathbf{dx} - \ell(\mathbf{v}) \, ,$$

where $\ell(\mathbf{v}) = \int_{\mathbf{\Omega}} \mathbf{f}\mathbf{v} \, \mathbf{dx} + \int_{\partial_{\mathbf{2}}\mathbf{\Omega}} \mathbf{F}\mathbf{v} \, \mathbf{ds}$.

In this formulation $(\mathbf{u}, \mathbf{p}) \in (\mathbf{V_0} + \mathbf{u_0}) \times \mathbf{Q}$ satisfies the relations

$$\int_\Omega \left(\mathbf{A}^{-1}\mathbf{p} - \nabla \mathbf{u}\right) \cdot \mathbf{q}\, \mathbf{dx} = \mathbf{0} \qquad \forall \mathbf{q} \in \mathbf{Q}\,, \tag{9.1}$$

$$\int_\Omega \mathbf{p} \cdot \nabla \mathbf{w}\, \mathbf{dx} - \ell(\mathbf{w}) = \mathbf{0} \qquad \forall \mathbf{w} \in \mathbf{V_0}\,. \tag{9.2}$$

Here

$$\begin{aligned} &\mathbf{p} = \mathbf{A}\nabla \mathbf{u}, \quad \text{is satisfied in } \mathbf{L_2}(\boldsymbol{\Omega}) - \text{sense} \\ &\mathbf{div p} + \mathbf{f} = \mathbf{0} \quad \text{in } \boldsymbol{\Omega} \text{ and} \\ &\mathbf{p} \cdot \mathbf{n} = \mathbf{F} \text{ on } \partial_2 \boldsymbol{\Omega} \quad \text{are satisfied in a weak sense.} \end{aligned}$$

As we have seen in previous lectures $\mathbf{L}$ generates two functionals

$$\mathbf{J}(\mathbf{v}) := \sup_{\mathbf{q}\in\mathbf{Q}} \mathbf{L}(\mathbf{v},\mathbf{q}) = \frac{1}{2} \| \nabla \mathbf{v} \|^2 - \ell(\mathbf{v})$$

and

$$\mathbf{I}^*(\mathbf{q}) := -\frac{1}{2} \| \mathbf{q} \|_*^2 - \ell(\mathbf{u_0}) + \int_{\Omega} \nabla \mathbf{u_0} \cdot \mathbf{q}\, \mathbf{dx}\,.$$

Also, we know that

$$\inf_{\mathbf{v}\in\mathbf{V_0}+\mathbf{u_0}} \mathbf{J}(\mathbf{v}) := \inf \mathcal{P} = \mathbf{L}(\mathbf{u},\mathbf{p}) = \sup \mathcal{P}^* := \sup_{\mathbf{q}\in\mathbf{Q}_\ell} \mathbf{I}^*(\mathbf{q})\,, \qquad (9.3)$$

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 (9.1)–(9.2) 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

$$\int\limits_{\Omega} \left(\mathbf{A}^{-1}\mathbf{p_h} - \nabla \mathbf{u_h}\right) \cdot \mathbf{q_h}\, \mathbf{dx} = \mathbf{0} \quad \forall \mathbf{q_h} \in \mathbf{Q_h}\,, \tag{9.4}$$

$$\int\limits_{\Omega} \mathbf{p_h} \cdot \nabla \mathbf{w_h}\, \mathbf{dx} - \ell(\mathbf{w_h}) = \mathbf{0} \quad \forall \mathbf{w_h} \in \mathbf{V_{0h}}\,. \tag{9.5}$$

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

$$\mathbf{g}(\mathbf{v}, \mathbf{q}) := \int\limits_{\Omega} (\nabla \mathbf{v} \cdot \mathbf{q} + \mathbf{v}(\mathbf{div q}))\, \mathbf{dx}\,. \tag{9.6}$$

We have

$$\mathbf{L}(\mathbf{v}, \mathbf{q}) = \int_{\Omega} \left(\nabla \mathbf{v} \cdot \mathbf{q} - \frac{1}{2} \mathbf{A}^{-1} \mathbf{q} \cdot \mathbf{q} \right) \mathbf{dx} - \ell(\mathbf{v}) =$$

$$= \mathbf{g}(\mathbf{v}, \mathbf{q}) - \int\limits_{\Omega} \mathbf{v}(\mathbf{div q})\, \mathbf{dx} - \frac{1}{2} \interleave \mathbf{q} \interleave_*^2 - \ell(\mathbf{v})\,.$$

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

$$\widehat{\mathbf{L}}(\mathbf{v},\mathbf{q}) :=$$

$$-\frac{\mathbf{1}}{\mathbf{2}} \Vert\, \mathbf{q} \,\Vert_*^{\mathbf{2}} - \int_{\Omega} \mathbf{v}(\mathbf{div}\mathbf{q})\,\mathbf{dx} - \int_{\Omega} \mathbf{f}\mathbf{v}\,\mathbf{dx} - \int_{\partial_2\Omega} \mathbf{F}\mathbf{u_0}\,\mathbf{ds} + \mathbf{g}(\mathbf{u_0},\mathbf{q})\,.$$

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

$$\widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{q}}) \;\leq\; \widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) \;\leq\; \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \widehat{\mathbf{p}}) \quad \forall \widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}\,,\ \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}} \tag{9.7}$$

lead to is the so-called

Dual Mixed Formulation

of the problem in question (see, e.g., F. Brezzi and M. Fortin).

From (9.7) 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

$$-\frac{1}{2} \,|\!|\!|\, \widehat{\mathbf{p}}+\lambda\boldsymbol{\eta} \,|\!|\!|_*^2 - \int\limits_{\Omega} \widehat{\mathbf{u}}(\mathbf{div}(\widehat{\mathbf{p}}+\lambda\boldsymbol{\eta})-\mathbf{f}\widehat{\mathbf{u}})\mathbf{dx} - \int\limits_{\partial_2\Omega} \mathbf{F}\mathbf{u_0}\,\mathbf{ds}+\mathbf{g}(\mathbf{u_0},\widehat{\mathbf{p}}+\lambda\boldsymbol{\eta}) \leq$$

$$-\frac{1}{2} \,|\!|\!|\, \widehat{\mathbf{p}} \,|\!|\!|_*^2 - \int\limits_{\Omega} \widehat{\mathbf{u}}(\mathbf{div}\widehat{\mathbf{p}})\,\mathbf{dx} - \int\limits_{\Omega} \mathbf{f}\widehat{\mathbf{u}}\,\mathbf{dx} - \int\limits_{\partial_2\Omega} \mathbf{F}\mathbf{u_0}\,\mathbf{ds} + \mathbf{g}(\mathbf{u_0},\widehat{\mathbf{p}}),$$

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

$$-\lambda \int\limits_{\Omega} (\mathbf{A}^{-1}\widehat{\mathbf{p}} \cdot \boldsymbol{\eta} + \widehat{\mathbf{u}}(\mathbf{div}\boldsymbol{\eta}))\mathbf{dx} + \lambda\mathbf{g}(\mathbf{u_0},\boldsymbol{\eta}) \leq \frac{\lambda^2}{2} \int_{\Omega} \mathbf{A}^{-1}\boldsymbol{\eta}\cdot\boldsymbol{\eta}\mathbf{dx}.$$

Rewrite it as

$$\int\limits_{\Omega}(\mathbf{A}^{-1}\widehat{\mathbf{p}}\cdot\boldsymbol{\eta}+\widehat{\mathbf{u}}(\mathbf{div}\boldsymbol{\eta}))\mathbf{dx}-\mathbf{g}(\mathbf{u_0},\boldsymbol{\eta})\geq\frac{\lambda}{\mathbf{2}}\int_{\Omega}\mathbf{A}^{-1}\boldsymbol{\eta}\cdot\boldsymbol{\eta}\mathbf{dx}.$$

Since $\lambda>0$ can be taken arbitrarily small, the latter relation may hold only if

$$\int\limits_{\Omega}(\mathbf{A}^{-1}\widehat{\mathbf{p}}\cdot\boldsymbol{\eta}+\widehat{\mathbf{u}}\mathbf{div}\boldsymbol{\eta})\mathbf{dx}-\mathbf{g}(\mathbf{u_0},\boldsymbol{\eta})\geq\mathbf{0}.$$

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

$$\int\limits_{\Omega}(\mathbf{A}^{-1}\widehat{\mathbf{p}}\cdot\boldsymbol{\eta}+\widehat{\mathbf{u}}\mathbf{div}\boldsymbol{\eta})\mathbf{dx}-\mathbf{g}(\mathbf{u_0},\boldsymbol{\eta})=\mathbf{0}\quad\forall\boldsymbol{\eta}\in\widehat{\mathbf{Q}}_{\mathbf{0}}.$$

From

$$\widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) \; \leq \; \widehat{\mathbf{L}}(\widehat{\mathbf{u}} + \widehat{\mathbf{v}}, \widehat{\mathbf{p}}) \quad \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}} := \mathbf{L}^2(\boldsymbol{\Omega})$$

we observe that the terms of $\widehat{\mathbf{L}}$ linear with respect to the "pressure" must vanish. Namely, we obtain

$$\int_{\boldsymbol{\Omega}} (\widehat{\mathbf{v}}\mathbf{div}\widehat{\mathbf{p}} + \mathbf{f}\widehat{\mathbf{v}})\mathbf{dx} = \mathbf{0}$$

Thus, we arrive at the system

$$\int_{\boldsymbol{\Omega}} \left(\mathbf{A}^{-1}\widehat{\mathbf{p}} \cdot \widehat{\mathbf{q}} + (\mathbf{div}\widehat{\mathbf{q}})\widehat{\mathbf{u}}\right) \mathbf{dx} = \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{q}}) \quad \forall \widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{0}}\,, \qquad (9.8)$$

$$\int_{\boldsymbol{\Omega}} (\mathbf{div}\widehat{\mathbf{p}} + \mathbf{f})\widehat{\mathbf{v}}\,\mathbf{dx} = \mathbf{0} \qquad \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}}\,. \qquad (9.9)$$

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* $\mathbf{F} \in \mathbf{L_2}(\partial_2\mathbf{\Omega})$ *there exists* $\mathbf{p^v} \in \widehat{\mathbf{Q}}_{\mathbf{F}}$ *such that*

$$\mathbf{divp^v} + \widehat{\mathbf{v}} = \mathbf{0} \ \text{ in } \mathbf{\Omega}\,, \tag{9.10}$$

$$\| \ \mathbf{p^v} \ \|_* \leq \mathbf{C_\Omega} \ (\|\widehat{\mathbf{v}}\| + \|\mathbf{F}\|_{\partial_2\mathbf{\Omega}}) \ . \tag{9.11}$$

Proof. We know that the boundary-value problem

$$\begin{aligned} \mathbf{divA\nabla u^v} + \widehat{\mathbf{v}} = \mathbf{0} \quad & \text{ in } \Omega\,, \\ \mathbf{u^v} = \mathbf{0} \quad & \text{on } \partial_1\Omega\,, \\ \mathbf{A\nabla u^v} \cdot \mathbf{n} = \mathbf{F} \quad & \text{ on } \partial_2\Omega \end{aligned}$$

possesses the unique solution $\mathbf{u^v} \in \mathbf{V_0}$.

For it and the energy estimate

$$\| \nabla \mathbf{u^v} \| \leq \mathbf{C_\Omega} \left(\|\widehat{\mathbf{v}}\| + \|\mathbf{F}\|_{\partial_2 \Omega} \right)$$

holds. Let $\mathbf{p^v} := \mathbf{A}\nabla\mathbf{u^v}$. We have

$$\mathbf{div p^v} + \widehat{\mathbf{v}} = \mathbf{0}.$$

Obviously, $\mathbf{p^v} \in \widehat{\mathbf{Q}}_\mathbf{F}$ and, since

$$\| \mathbf{p^v} \|_*^2 = \int\limits_\Omega \mathbf{A}^{-1}(\mathbf{A}\nabla\mathbf{u^v}) \cdot (\mathbf{A}\nabla\mathbf{u^v})\, \mathbf{dx} = \| \nabla\mathbf{u^v} \|^2,$$

we find that (9.11) also holds.
□

By the Lemma we can easily prove the coercivity of $\widehat{\mathbf{J}}$ on $\widehat{\mathbf{V}}$. Indeed,

$$\widehat{\mathbf{J}}(\widehat{\mathbf{v}}) \geq \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \alpha \mathbf{p}^{\mathbf{v}}) =$$

$$-\frac{1}{2} \parallel\!| \alpha \mathbf{p}^{\mathbf{v}} \parallel\!|_*^2 - \alpha \int\limits_{\Omega} \widehat{\mathbf{v}}(\mathbf{div} \mathbf{p}^{\mathbf{v}})\, \mathbf{dx} - \int\limits_{\Omega} \mathbf{f} \widehat{\mathbf{v}}\, \mathbf{dx} - \int\limits_{\partial_2 \Omega} \mathbf{F} \mathbf{u}_0\, \mathbf{ds} + \mathbf{g}(\mathbf{u}_0, \alpha \mathbf{p}^{\mathbf{v}}) =$$

$$= -\frac{1}{2} \alpha^2 \parallel\!| \mathbf{p}^{\mathbf{v}} \parallel\!|_*^2 + \alpha \|\widehat{\mathbf{v}}\|^2 - \|\mathbf{f}\| \|\widehat{\mathbf{v}}\| + \mathbf{g}(\mathbf{u}_0, \alpha \mathbf{p}^{\mathbf{v}}) - \int\limits_{\partial_2 \Omega} \mathbf{F} \mathbf{u}_0\, \mathbf{ds}\,.$$

Here $|\mathbf{g}(\mathbf{u}_0, \alpha \mathbf{p}^{\mathbf{v}})| \leq \alpha \|\mathbf{p}^{\mathbf{v}}\|_{\mathbf{div}} \|\mathbf{u}_0\|_{1,2,\Omega}$ and

$$\|\mathbf{p}^{\mathbf{v}}\|_{\mathbf{div}}^2 = \|\mathbf{p}^{\mathbf{v}}\|^2 + \|\mathbf{div} \mathbf{p}^{\mathbf{v}}\|^2 \leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_1} \parallel\!| \mathbf{p}^{\mathbf{v}} \parallel\!|_*^2 + \|\widehat{\mathbf{v}}\|^2 \leq$$

$$\leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_1} \mathbf{C}_{\Omega}^2 \left(\|\widehat{\mathbf{v}}\| + \|\mathbf{F}\|_{\partial_2 \Omega}\right)^2 + \|\widehat{\mathbf{v}}\|^2\,.$$

Therefore

$$\widehat{\mathbf{J}}(\widehat{\mathbf{v}}) \geq -\frac{\mathbf{1}}{\mathbf{2}}\alpha^{\mathbf{2}}\mathbf{C}_{\boldsymbol{\Omega}}^{\mathbf{2}}\|\widehat{\mathbf{v}}\|^{\mathbf{2}} + \alpha\|\widehat{\mathbf{v}}\|^{\mathbf{2}} + \boldsymbol{\Theta}(\|\widehat{\mathbf{v}}\|) + \boldsymbol{\Theta}_{\mathbf{0}}\,,$$

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 $\alpha = \mathbf{1}/\mathbf{C}_{\boldsymbol{\Omega}}^{\mathbf{2}}$. Then

$$\widehat{\mathbf{J}}(\widehat{\mathbf{v}}) \geq \frac{\mathbf{1}}{\mathbf{2C}_{\boldsymbol{\Omega}}^{\mathbf{2}}}\|\widehat{\mathbf{v}}\|^{\mathbf{2}} + \boldsymbol{\Theta}(\|\widehat{\mathbf{v}}\|) + \boldsymbol{\Theta}_{\mathbf{0}} \longrightarrow +\infty \text{ as } \|\widehat{\mathbf{v}}\| \rightarrow \infty\,.$$

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

$$\inf_{\substack{\phi \in \mathbf{L}^2(\mathbf{\Omega}) \\ \psi \in \mathbf{L}^2(\partial_2 \mathbf{\Omega})}} \sup_{\mathbf{q} \in \widehat{\mathbf{Q}}_\mathbf{F}} \frac{\int\limits_{\mathbf{\Omega}} \phi \, \mathbf{div} \mathbf{q} \, \mathbf{dx} + \int_{\partial_2 \mathbf{\Omega}} \psi \mathbf{q} \cdot \mathbf{n} \, \mathbf{ds}}{\|\mathbf{q}\|_{\mathbf{div}} (\|\phi\|^2 + \|\psi\|^2_{\partial_2 \mathbf{\Omega}})^{1/2}} \geq \mathbf{C_0} > \mathbf{0} \, .$$

The Dual Problem with respect to the Lagrangian $\widehat{\mathbf{L}}$

Let us now construct the dual functional $\widehat{\mathbf{l}}^*$. It is easy to see that

$$\widehat{\mathbf{l}}^*(\widehat{\mathbf{q}}) = \inf_{\widehat{\mathbf{v}}} \widehat{\mathbf{L}}(\widehat{\mathbf{v}}, \widehat{\mathbf{q}}) =$$

$$= \inf_{\widehat{\mathbf{v}}} \left\{ -\frac{1}{2} \| \widehat{\mathbf{q}} \|_*^2 - \int_\Omega \mathbf{v}(\mathbf{div}\widehat{\mathbf{q}})\mathbf{dx} - \int_\Omega \mathbf{fvdx} - \int_{\partial_2\Omega} \mathbf{Fu_0 ds} + \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{q}}) \right\} =$$

$$= -\frac{1}{2} \| \widehat{\mathbf{q}} \|_*^2 + \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{q}}) - \int_{\partial_2\Omega} \mathbf{Fu_0}\, \mathbf{ds}$$

provided that $\mathbf{div}\widehat{\mathbf{q}} + \mathbf{f} = \mathbf{0}$ (in the $\mathbf{L_2}$-sense). In all other cases $\widehat{\mathbf{l}}^*(\widehat{\mathbf{q}}) = -\infty$.

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

$$\widehat{\mathbf{I}}^*(\mathbf{q}) = -\frac{1}{2} \mathrel{|\!|\!|} \widehat{\mathbf{q}} \mathrel{|\!|\!|}_*^2 + \int\limits_{\Omega} (\nabla \mathbf{u_0} \cdot \widehat{\mathbf{q}} - \mathbf{f u_0})\, \mathbf{dx} - \int_{\partial_2 \Omega} \mathbf{F u_0}\, \mathbf{ds}$$

$$= \int\limits_{\Omega} \nabla \mathbf{u_0} \cdot \widehat{\mathbf{q}} \mathbf{dx} - \frac{1}{2} \mathrel{|\!|\!|} \widehat{\mathbf{q}} \mathrel{|\!|\!|}_*^2 - \ell(\mathbf{u_0})\,,$$

Since $\widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{F}}$, we have

$$\int\limits_{\Omega} \nabla \mathbf{w} \cdot \widehat{\mathbf{q}} \mathbf{dx} = -\int\limits_{\Omega} (\mathbf{div}\widehat{\mathbf{q}}) \mathbf{w dx} + \int_{\partial_2 \Omega} \mathbf{F w ds} \quad \forall \mathbf{w} \in \mathbf{V_0}\,.$$

we see that $\widehat{\mathbf{q}}$ satisfies the relation

$$\int\limits_{\Omega} \nabla \mathbf{w} \cdot \widehat{\mathbf{q}} \mathbf{dx} = \ell(\mathbf{w}) \quad \forall \mathbf{w} \in \mathbf{V_0}\,.$$

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 $\mathbf{I}^*$ on the set Q_ℓ. This means that

$$\sup \mathcal{P}^* = \sup \widehat{\mathcal{P}}^*.$$

Since the saddle point of $\widehat{\mathbf{L}}$ exists, we have

$$\widehat{\mathbf{L}}(\widehat{\mathbf{u}}, \widehat{\mathbf{p}}) = \inf \widehat{\mathcal{P}} = \sup \widehat{\mathcal{P}}^*,$$

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

$$\int\limits_{\Omega} \nabla \mathbf{w} \cdot \mathbf{p}\, \mathbf{dx} = \int\limits_{\Omega} \mathbf{fw}\, \mathbf{dx} + \int_{\partial_2 \Omega} \mathbf{Fw}\, \mathbf{ds} \quad \forall \mathbf{w} \in \mathbf{V_0}\,.$$

From here we see that $\mathbf{divp} + \mathbf{f} = \mathbf{0}$ a.e. in Ω and, hence,

$$\int\limits_{\Omega} (\nabla \mathbf{w} \cdot \mathbf{p} + (\mathbf{divp})\mathbf{w})\, \mathbf{dx} = \int_{\partial_2 \Omega} \mathbf{Fw}\, \mathbf{ds} \quad \forall \mathbf{w} \in \mathbf{V_0}\,,$$

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

$$\int_\Omega \left(\mathbf{A}^{-1}\widehat{\mathbf{p}}_\mathbf{h} \cdot \widehat{\mathbf{q}}_\mathbf{h} + \widehat{\mathbf{u}}_\mathbf{h}\mathbf{div}\widehat{\mathbf{q}}_\mathbf{h}\right)\mathbf{dx} = \mathbf{g}(\mathbf{u_0}, \widehat{\mathbf{q}}_\mathbf{h}) \quad \forall \widehat{\mathbf{q}}_\mathbf{h} \in \widehat{\mathbf{Q}}_{\mathbf{0h}}, \tag{9.12}$$

$$\int_\Omega (\mathbf{div}\widehat{\mathbf{p}}_\mathbf{h} + \mathbf{f})\widehat{\mathbf{v}}_\mathbf{h}\mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}}_\mathbf{h} \in \widehat{\mathbf{V}}_\mathbf{h}. \tag{9.13}$$

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 **A**. In this case, the basic system is as follows

$$\int_{\Omega} \left(\mathbf{A}^{-1}\widehat{\mathbf{p}} \cdot \widehat{\mathbf{q}} + (\mathbf{div}\widehat{\mathbf{q}})\widehat{\mathbf{u}}\right)\, \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}} \in \widehat{\mathbf{Q}}_{\mathbf{0}}\,,$$

$$\int_{\Omega} (\mathbf{div}\widehat{\mathbf{p}} + \mathbf{f})\widehat{\mathbf{v}}\, \mathbf{dx} = \mathbf{0} \qquad \forall \widehat{\mathbf{v}} \in \widehat{\mathbf{V}}\,.$$

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

$$\int_{\Omega} \left(\mathbf{A}^{-1} \widehat{\mathbf{p}}_{\mathbf{h}} \cdot \widehat{\mathbf{q}}_{\mathbf{h}} + \widehat{\mathbf{u}}_{\mathbf{h}} \mathbf{div} \widehat{\mathbf{q}}_{\mathbf{h}} \right) \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{h}},$$

$$\int_{\Omega} (\mathbf{div} \widehat{\mathbf{p}}_{\mathbf{h}} + \mathbf{f}) \widehat{\mathbf{v}}_{\mathbf{h}} \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}}_{\mathbf{h}} \in \widehat{\mathbf{V}}_{\mathbf{h}}.$$

Assumptions.
(a) $\mathcal{T}_h$ is a regular triangulation of a polygonal domain Ω.
(b) $\widehat{\mathbf{V}}_h = \{\mathbf{v}_\mathbf{h} \in \mathbf{L}^2 \,|\, \mathbf{v}_\mathbf{h} \in \mathbf{P}^0(\mathbf{T}) \ \forall \mathbf{T} \in \mathcal{T}_\mathbf{h}\}$.
(c) $\widehat{\mathbf{Q}}_h = \{\mathbf{q}_\mathbf{h} \in \mathbf{H}(\Omega, \mathbf{div}) \,|\, \mathbf{q}_\mathbf{h} \in \mathbf{RT}^0(\mathbf{T}) \ \forall \mathbf{T} \in \mathcal{T}_\mathbf{h}\}$.
(d) $\mathbf{f} \in \mathbf{P}^0(\mathbf{T}), \quad \forall \mathbf{T} \in \mathcal{T}_\mathbf{h}$

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

We need that one more condition be satisfied in order to provide the stability of the discrete DM formulation.
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 $\mathbf{q_h^v} \in \widehat{\mathbf{Q}}_\mathbf{h}$ such that

$$\mathbf{div q_h^v} = \mathbf{v_h} \quad \text{(compatibility)}, \tag{9.14}$$

$$\|\mathbf{q_h^v}\| \leq \mathbf{C}\|\mathbf{v_h}\| \quad \text{(stability)}. \tag{9.15}$$

Discrete Inf-Sup condition

From (9.14) and (9.15), it follows that

$$\inf_{\mathbf{v_h} \in \widehat{\mathbf{V}}_\mathbf{h}} \sup_{\mathbf{q_h} \in \widehat{\mathbf{Q}}_\mathbf{h}} \frac{\int_\Omega \mathbf{v_h} \mathbf{div} \mathbf{q_h} \, \mathbf{dx}}{\|\mathbf{v_h}\| \|\mathbf{q_h}\|_{\mathbf{div}}} \geq \mathbb{C} > \mathbf{0}$$

Indeed,

$$\sup_{\mathbf{q_h} \in \widehat{\mathbf{Q}}_\mathbf{h}} \frac{\int_\Omega \mathbf{v_h} \mathbf{div} \mathbf{q_h} \, \mathbf{dx}}{\|\mathbf{v_h}\| \|\mathbf{q_h}\|_{\mathbf{div}}} \geq \frac{\int_\Omega \mathbf{v_h} \mathbf{div} \mathbf{q_h^v} \, \mathbf{dx}}{\|\mathbf{v_h}\| \|\mathbf{q_h^v}\|_{\mathbf{div}}} = \frac{\|\mathbf{v_h}\|}{\|\mathbf{q_h}\|_{\mathbf{div}}} \geq \frac{\mathbf{1}}{\sqrt{\mathbf{1} + \mathbf{C}^2}}.$$

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

$$\int_\Omega \mathbf{A}^{-1}\mathbf{p} \cdot \mathbf{q} \, d\mathbf{x} = \mathbf{0} \qquad \forall \mathbf{q} \in \mathbf{Q_0},$$

where $\mathbf{Q_0}$ is the space of solenoidal functions. Therefore, for any $\mathbf{q} \in \mathbf{Q_f}$,

$$\frac{1}{2} \| \mathbf{q} - \mathbf{p} \|_*^2 = \frac{1}{2} \| \mathbf{q} \|_*^2 - \frac{1}{2} \| \mathbf{p} \|_*^2 + \int_\Omega \mathbf{A}^{-1}\mathbf{p} \cdot (\mathbf{p} - \mathbf{q}) d\mathbf{x} =$$
$$= \frac{1}{2} \| \mathbf{q} \|_*^2 - \frac{1}{2} \| \mathbf{p} \|_*^2 .$$

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

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \inf_{\mathbf{q_{fh}} \in \mathbf{Q_{fh}}} \| \mathbf{p} - \mathbf{q_{fh}} \|_* \,. \tag{9.16}$$

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

$$\int\limits_{\Omega} \left(\mathbf{A}^{-1}\widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}} \cdot \widehat{\mathbf{q}}_{\mathbf{h}} + \widehat{\mathbf{u}}^{\mathbf{f}}_{\mathbf{h}} \mathbf{div} \widehat{\mathbf{q}}_{\mathbf{h}}\right) \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{0h}}, \tag{9.17}$$

$$\int\limits_{\Omega} (\mathbf{div}\widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}} + \widetilde{\mathbf{f}})\widehat{\mathbf{v}}_{\mathbf{h}} \mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{v}}_{\mathbf{h}} \in \widehat{\mathbf{V}}_{\mathbf{h}}. \tag{9.18}$$

Under the assumptions made $\widetilde{\mathbf{f}} \in \mathbf{P}^{\mathbf{0}}(\mathbf{T})$, 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

$$\bar{\mathbf{c}}_1 \|\widehat{\mathbf{p}}_h^f\|^2 \leq \|\!| \widehat{\mathbf{p}}_h^f \|\!|_*^2 \leq \|\widehat{\mathbf{u}}_h^f\| \|\widetilde{\mathbf{f}}\|. \tag{9.19}$$

By (9.14) and (9.15) we conclude that for $\widehat{\mathbf{u}}_h^f$ we can find $\bar{\mathbf{q}}_h$ in $\widehat{\mathbf{Q}}_h$ such that

$$\mathbf{div}\bar{\mathbf{q}}_h + \widehat{\mathbf{u}}_h^f = \mathbf{0} \quad \text{and} \quad \|\bar{\mathbf{q}}_h\| \leq \mathbf{C}\|\widehat{\mathbf{u}}_h^f\|$$

Use $\bar{\mathbf{q}}_h$ in the first identity (9.17). We have,

$$\int_\Omega \left(\mathbf{A}^{-1}\widehat{\mathbf{p}}_h^f \cdot \bar{\mathbf{q}}_h + \widehat{\mathbf{u}}_h^f \mathbf{div}\bar{\mathbf{q}}_h\right)\mathbf{dx} = \mathbf{0}$$

Thus,

$$\|\widehat{\mathbf{u}}_h^f\|^2 = \int_\Omega \widehat{\mathbf{u}}_h^f \mathbf{div}\bar{\mathbf{q}}_h \leq \|\!| \widehat{\mathbf{p}}_h^f \|\!|_* \|\!| \bar{\mathbf{q}}_h \|\!|_* \leq$$

$$\leq \bar{\mathbf{c}}_2 \|\!| \widehat{\mathbf{p}}_h^f \|\!|_* \|\bar{\mathbf{q}}_h\| \leq \bar{\mathbf{c}}_2 \mathbf{C} \|\!| \widehat{\mathbf{p}}_h^f \|\!|_* \|\widehat{\mathbf{u}}_h^f\|.$$

We observe that

$$\|\widehat{\mathbf{u}}^{\mathbf{f}}_{\mathbf{h}}\| \leq \bar{\mathbf{c}}_2 \mathbf{C} \; \|\!| \; \widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}} \; \|\!|_* \; . \tag{9.20}$$

Now, we use (9.19) and obtain

$$\|\!| \; \widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}} \; \|\!|_*^2 \leq \|\widehat{\mathbf{u}}^{\mathbf{f}}_{\mathbf{h}}\| \|\widetilde{\mathbf{f}}\| \leq \bar{\mathbf{c}}_2 \mathbf{C} \; \|\!| \; \widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}} \; \|\!|_* \; \|\widetilde{\mathbf{f}}\|.$$

so that

$$\|\!| \; \widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}} \; \|\!|_* \leq \bar{\mathbf{c}}_2 \mathbf{C} \|\widetilde{\mathbf{f}}\|. \tag{9.21}$$

Hence,

$$\|\widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}}\|^2 = \|\widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}}\|^2 + \|\mathbf{div}\widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}}\|^2 \leq \left(\mathbf{1} + \frac{\mathbf{c}_2^2}{\mathbf{c}_1^2}\mathbf{C}^2\right)\|\widetilde{\mathbf{f}}\|^2. \tag{9.22}$$

We note that the estimates (9.20), (9.21), and (9.22) show that the modified DM problem is **stable**, i.e. its solutions $(\widehat{\mathbf{p}}^{\mathbf{f}}_{\mathbf{h}}, \widehat{\mathbf{u}}^{\mathbf{f}}_{\mathbf{h}})$ 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 $\boldsymbol{\eta}_h = \widehat{\mathbf{p}}_h^f + \widehat{\mathbf{q}}_h$, where $\widehat{\mathbf{q}}_h$ is *an arbitrary element of* $\widehat{\mathbf{Q}}_\mathbf{h}$.
We have,

$$\begin{aligned} \mathbf{div}\eta_\mathbf{h} = \mathbf{div}\widehat{\mathbf{p}}_\mathbf{h}^\mathbf{f} + \mathbf{div}\widehat{\mathbf{q}}_\mathbf{h} = -\widetilde{\mathbf{f}} + \mathbf{div}\widehat{\mathbf{q}}_\mathbf{h} = \\ = \mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h}) + \mathbf{div}\widehat{\mathbf{q}}_\mathbf{h} = \mathbf{divp} = -\mathbf{f}. \end{aligned}$$

Therefore, $\eta_\mathbf{h} \in \mathbf{Q_f}$

Now, we recall the projection inequality and substitute in it η_h:

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \| \mathbf{p} - \eta_\mathbf{h} \|_* = \| \mathbf{p} - \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} - \widehat{\mathbf{q}}_\mathbf{h} \|_* \leq$$
$$\leq \| \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \|_* + \| \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \|_*$$

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, by means of (9.21) we obtain

$$\|\mathbf{p} - \mathbf{p_h}\|_{\mathbf{div}} \leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_\mathbf{1}} (\| \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \|_* + \| \widehat{\mathbf{p}}^\mathbf{f}_\mathbf{h} \|_*)$$
$$\leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_\mathbf{1}} (\| \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \|_* + \bar{\mathbf{c}}_\mathbf{2} \mathbf{C} \|\widetilde{\mathbf{f}}\|).$$

But $\widetilde{\mathbf{f}} = \mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h})$.

Thus, we arrive at the estimate

$$\|\mathbf{p} - \mathbf{p_h}\|_{\mathbf{div}} \leq$$
$$\leq \frac{\mathbf{1}}{\bar{\mathbf{c}}_{\mathbf{1}}}\left(\|\!|\, \mathbf{p} - \widehat{\mathbf{q}}_{\mathbf{h}} \,|\!\|_* + \bar{\mathbf{c}}_{\mathbf{2}}\mathbf{C}\|\mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_{\mathbf{h}})\|\right) \quad \forall \widehat{\mathbf{q}}_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{h}}.$$

and, therefore,

$$\|\mathbf{p} - \mathbf{p_h}\|_{\mathbf{div}} \leq \bar{\mathbf{C}}_{\mathbf{p}} \inf_{\widehat{\mathbf{q}}_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{h}}} \left\{\|\!|\, \mathbf{p} - \widehat{\mathbf{q}}_{\mathbf{h}} \,|\!\|_* + \|\mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_{\mathbf{h}})\|\right\}. \quad (9.23)$$

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

$$[\mathbf{u}]_{\mathbf{T}} = \frac{1}{|\mathbf{T}|}\int_{\mathbf{T}} \mathbf{u}d\mathbf{x}, \qquad [\mathbf{u}]_{\mathbf{h}}(\mathbf{x}) = [\mathbf{u}]_{\mathbf{T_i}} \quad \text{if} \quad \mathbf{x} \in \mathbf{T_i}.$$

Since $\mathbf{div}\widehat{\mathbf{q}}_{\mathbf{h}}$ is constant on each $\mathbf{T_i}$, we rewrite the relation as follows:

$$\int\limits_{\Omega}\Big(\mathbf{A}^{-1}(\widehat{\mathbf{p}}_{\mathbf{h}} - \mathbf{p}) \cdot \widehat{\mathbf{q}}_{\mathbf{h}} + (\widehat{\mathbf{u}}_{\mathbf{h}} - [\mathbf{u}]_{\mathbf{h}})\mathbf{div}\widehat{\mathbf{q}}_{\mathbf{h}}\Big)\mathbf{dx} = \mathbf{0} \quad \forall \widehat{\mathbf{q}}_{\mathbf{h}} \in \widehat{\mathbf{Q}}_{\mathbf{h}}.$$

Note that $[\mathbf{u}]_{\mathbf{h}} \in \widehat{\mathbf{V}}_{\mathbf{h}}$ and $\bar{\mathbf{u}}_{\mathbf{h}} := \widehat{\mathbf{u}}_{\mathbf{h}} - [\mathbf{u}]_{\mathbf{h}} \in \widehat{\mathbf{V}}_{\mathbf{h}}$ Now, we exploit the compatibility and stability conditions (9.14) and (9.15) 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 $\mathbf{q}'_h$ in the integral relation. We have

$$\int_\Omega \left(\mathbf{A}^{-1}(\widehat{\mathbf{p}}_h - \mathbf{p}) \cdot \mathbf{q}'_h + \bar{\mathbf{u}}_h \mathbf{div} \mathbf{q}'_h\right) \mathbf{dx} = \mathbf{0}.$$

From here, we conclude that

$$\begin{aligned} \|\bar{\mathbf{u}}_h\|^2 &= \left| \int_\Omega \mathbf{A}^{-1}(\widehat{\mathbf{p}}_h - \mathbf{p}) \cdot \mathbf{q}'_h \right| \leq \\ &\leq \| \widehat{\mathbf{p}}_h - \mathbf{p} \|_* \| \mathbf{q}'_h \|_* \leq \mathbf{C}\, \bar{\mathbf{c}}_2 \| \widehat{\mathbf{p}}_h - \mathbf{p} \|_* \, \|\bar{\mathbf{u}}_h\|. \end{aligned}$$

Thus,

$$\|\bar{\mathbf{u}}_h\| = \|[\mathbf{u}]_h - \widehat{\mathbf{u}}_h\| \ \leq \ \mathbf{C}\, \bar{\mathbf{c}}_2 \| \widehat{\mathbf{p}}_h - \mathbf{p} \|_* \ .$$

Since

$$\|\mathbf{u} - \widehat{\mathbf{u}}_h\| \leq \|\mathbf{u} - [\mathbf{u}]_h\| + \|[\mathbf{u}]_h - \widehat{\mathbf{u}}_h\| \leq$$
$$\leq \|\mathbf{u} - [\mathbf{u}]_h\| + \mathbf{C}\,\bar{\mathbf{c}}_2 \, ||| \, \widehat{\mathbf{p}}_h - \mathbf{p} \, |||_*$$

Note that by the definition of $[\mathbf{u}]_h$

$$\|\mathbf{u} - [\mathbf{u}]_h\| \leq \|\mathbf{u} - \mathbf{v}_h\| \qquad \forall \mathbf{v}_h \in \widehat{\mathbf{V}}_h.$$

From here, we observe that

$$\|\mathbf{u} - \widehat{\mathbf{u}}_h\| \leq \mathbf{C}\,\bar{\mathbf{c}}_2 \, ||| \, \widehat{\mathbf{p}}_h - \mathbf{p} \, |||_* + \inf_{\mathbf{v}_h \in \widehat{\mathbf{V}}_h} \|\mathbf{u} - \mathbf{v}_h\|$$

Recall that

$$||| \, \mathbf{p} - \mathbf{p}_h \, |||_* \leq ||| \, \mathbf{p} - \widehat{\mathbf{q}}_h \, |||_* + ||| \, \widehat{\mathbf{p}}^f_h \, |||_* \leq$$
$$||| \, \mathbf{p} - \widehat{\mathbf{q}}_h \, |||_* + \bar{\mathbf{c}}_2 \, \mathbf{C} \, \|\mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_h)\|.$$

Then, we arrive at the projection type error estimate for the primal variable

$$\|\mathbf{u} - \widehat{\mathbf{u}}_\mathbf{h}\| \leq$$
$$\leq \mathbf{C_u} \inf_{\widehat{\mathbf{q}}_\mathbf{h} \in \widehat{\mathbf{Q}}_\mathbf{h}} \Big\{ \| \mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h} \|_* + \|\mathbf{div}(\mathbf{p} - \widehat{\mathbf{q}}_\mathbf{h})\| +$$
$$+ \inf_{\mathbf{v_h} \in \widehat{\mathbf{V}}_\mathbf{h}} \|\mathbf{u} - \mathbf{v_h}\| \Big\}, \quad (9.24)$$

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 (9.23) and (9.25) lead to a qualified a priori convergence estimates provided that the solution possesses proper regularity.

A posteriori estimates for the primal mixed formulation

Further analysis follows the lines of the paper
S. Repin and A. Smolianski, A Functional-type a posteriori error estimates for mixed finite element methods. Russian J. Numer. Anal. Math. Modelling 20 (2005), no. 4, 365–382.
A posteriori estimates for the mixed formulation are based on the relation that we have already derived:

$$\|\!| \mathbf{p}-\mathbf{q} \|\!|_*^2 + \|\!| \nabla(\mathbf{u}-\mathbf{v}) \|\!|^2 = 2(\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q}))\,, \tag{9.25}$$

where $\mathbf{q} \in \mathbf{Q}_\ell$ 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

$$\begin{aligned} \| \mathbf{p} - \mathbf{q} \|_*^2 + \| \nabla(\mathbf{u} - \mathbf{v}) \|^2 \leq \mathbf{2}(\mathbf{1} + \beta)\mathbf{D}(\nabla \mathbf{v}, \mathbf{y}) \\ + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \mathbf{C}^2 \left(\|\mathbf{div y} + \mathbf{f}\|^2 + \|\mathbf{y} \cdot \mathbf{n} - \mathbf{F}\|^2_{\partial_2 \Omega} \right), \end{aligned} \quad (9.26)$$

where $\mathbf{y} \in \widehat{\mathbf{Q}}^+$, $\mathbf{q} \in \mathbf{Q}_\ell$ and $v \in V_0 + u_0$ are arbitrary functions and β is any positive number.

Thus, for the error in the primal variable we have

$$\begin{aligned} \| \nabla(\mathbf{u} - \mathbf{u_h}) \|^2 \leq & \; \mathbf{2}(\mathbf{1} + \beta)\mathbf{D}(\nabla \mathbf{u_h}, \mathbf{y}) \\ & + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \mathbf{C}^2 \left(\|\mathbf{div y} + \mathbf{f}\|^2 + \|\mathbf{y} \cdot \mathbf{n} - \mathbf{F}\|^2_{\partial_2 \Omega} \right) . \end{aligned} \quad (9.27)$$

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

By using the general estimate derived in Lecture 4, we find that

$$\| \mathbf{p} - \mathbf{p_h} \|_* \leq \sqrt{2}\mathbf{D}^{1/2}(\nabla \mathbf{v}, \mathbf{y}) + \| \mathbf{y} - \mathbf{p_h} \|_* \\ + \mathbf{2C}\left(\|\mathbf{div y} + \mathbf{f}\|^2 + \|\mathbf{y} \cdot \mathbf{n} - \mathbf{F}\|^2 \right)^{1/2} . \quad (9.28)$$

Here $\mathbf{v}$ 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 $v = u$, then the right-hand side of (9.37) 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 (9.37), 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:

$$\widetilde{\mathbf{q}}_{\mathbf{ij}} \cdot \mathbf{n}_{\mathbf{ij}} = \frac{1}{2}(\mathbf{q}_{\mathbf{i}} + \mathbf{q}_{\mathbf{j}}) \cdot \mathbf{n}_{\mathbf{ij}}\,,$$

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

$$\widetilde{\mathbf{q}}_{\mathbf{i0}} \cdot \mathbf{n}_{\mathbf{i0}} = \frac{1}{|\mathbf{E}_{\mathbf{i0}}|}\int_{\mathbf{E}_{\mathbf{i0}}} \mathbf{F}\,\mathbf{ds}\,.$$

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

$$\frac{\mathbf{1}}{\mathbf{2}} \, \|\, \mathbf{y} - \mathbf{p} \,\|_*^{\mathbf{2}} \leq (1+\gamma)\left(1 + \frac{1}{\gamma} + \frac{1}{\beta\gamma}\right) [\![\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,]\!]^{\mathbf{2}} + \\ +(\mathbf{1}+\beta)\left(\mathbf{1} + \frac{\mathbf{1}}{\gamma}\right) \mathbf{D}(\mathbf{\Lambda v}, \mathbf{y}).$$

that we have derived in Lecture 5. 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 $[\![\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,]\!]$ from above in the same way we did it in Lecture 6. Minimization with respect to γ and β leads to the estimate

$$\| \mathbf{p} - \widehat{\mathbf{p}}_{\mathbf{h}} \|_* \leq \sqrt{2}\mathbf{D}^{1/2}(\nabla \mathbf{v}, \widehat{\mathbf{p}}_{\mathbf{h}}) + \tag{9.29}$$
$$+ \mathbf{2C} \left(\| \mathrm{\mathbf{div}} \widehat{\mathbf{p}}_{\mathbf{h}} + \mathbf{f} \|^{\mathbf{2}} + \| \widehat{\mathbf{p}}_{\mathbf{h}} \cdot \mathbf{n} - \mathbf{F} \|^{\mathbf{2}}_{\partial_{\mathbf{2}} \mathbf{\Omega}} \right)^{\mathbf{1/2}},$$

where $\mathbf{v}$ is an arbitrary function from $\mathbf{V_0} + \mathbf{u_0}$.

For the sake of simplicity we assume that Ω is a polygonal domain decomposed into a regular collection of simplexes. If $\widehat{p}_h$ is constructed by means of RT_0-elements, then

$$\int\limits_{\Omega} (\mathrm{div}\widehat{p}_h + f) w_h \, dx = 0 \quad \forall w_h \in \widehat{V}_h \subset \widehat{V}\,, \tag{9.30}$$

where the subspace $\widehat{V}_h$ contains piecewise constant functions. Therefore, on each element T_i

$$\mathrm{div}\widehat{p}_h = -\frac{1}{|T_i|} \int_{T_i} f \, dx\,. \tag{9.31}$$

Let us define by $[f]$ the function that belongs to $\widehat{V}_h$ and whose values on T_i coincide with the mean values of f on T_i. Then, we have

$$\mathrm{div}\widehat{p}_h = -[f] \quad \text{on every} \quad T_i.$$

Remark. We observe that estimate (9.30) is valid for any approximate flux $\widehat{\mathbf{p}}_{\mathbf{h}}$ from $\widehat{\mathbf{Q}}^{+}$. If $\widehat{\mathbf{p}}_{\mathbf{h}}$ were in the narrower set $\widehat{\mathbf{Q}}_{\mathbf{F}}$ (as it is supposed to be in the discrete dual mixed method) the last norm in (9.30) 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 $\partial_{\mathbf{2}}\mathbf{\Omega}$.

The problem of taking into account the essential boundary condition for the flux variable

$$\widehat{\mathbf{p}} \cdot \mathbf{n} = \mathbf{F} \quad \text{on} \quad \partial_2 \boldsymbol{\Omega}$$

in the dual mixed method is not 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. Partial Diff. Eq. **19**(2) (2003), 192–210).

Since (9.30) still works for such approximations of the flux, we propose a simple modification of the discrete dual method, particularly suited for the lowest-order Raviart-Thomas approximation.

Namely, instead of requiring $\widehat{\mathbf{p}}_\mathbf{h} \in \widehat{\mathbf{Q}}_\mathbf{F}$, we impose a weaker condition

$$\widehat{\mathbf{p}}_\mathbf{h} \cdot \mathbf{n}\Big|_{\mathbf{E_{i0}}} = \frac{\mathbf{1}}{|\mathbf{E_{i0}}|} \int_{\mathbf{E_{i0}}} \mathbf{F}\, \mathbf{ds} \tag{9.32}$$

on every edge $\mathbf{E_{i0}} \in \partial_\mathbf{2}\boldsymbol{\Omega}$. The space of test functions $\widehat{\mathbf{Q}}_\mathbf{0h} \subset \widehat{\mathbf{Q}}_\mathbf{0}$ will obviously consist of the $\mathbf{RT_0}$-approximations $\widehat{\mathbf{q}}_\mathbf{h}$ such that $\widehat{\mathbf{q}}_\mathbf{h} \cdot \mathbf{n} = \mathbf{0}$ on each edge $\mathbf{E_{i0}} \in \partial_\mathbf{2}\boldsymbol{\Omega}$.

If now we denote by $[\mathbf{F}]$ the piecewise constant function defined on the set of edges forming $\partial_\mathbf{2}\boldsymbol{\Omega}$ and whose value on every edge $\mathbf{E_{i0}} \in \partial_\mathbf{2}\boldsymbol{\Omega}$ 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 $\mathbf{E_{i0}} \in \partial_\mathbf{2}\boldsymbol{\Omega}$.

As a result, we obtain from (9.30)

$$|\!|\!|\ \mathbf{p} - \widehat{\mathbf{p}}_\mathbf{h}\ |\!|\!|_* \leq \sqrt{\mathbf{2}}\mathbf{D}^{\mathbf{1/2}}(\nabla \mathbf{v}, \widehat{\mathbf{p}}_\mathbf{h}) + \mathbf{2C}\left(\|\mathbf{f} - [\mathbf{f}]\|^\mathbf{2} + \|\mathbf{F} - [\mathbf{F}]\|^\mathbf{2}_{\partial_\mathbf{2}\boldsymbol{\Omega}}\right)^{\mathbf{1/2}}. \tag{9.33}$$

The question that now arises is how to choose in (9.33) the function $\mathbf{v} \in \mathbf{V_0} + \mathbf{u_0}$. 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 $\mathbf{w_h} \in \mathbf{V_0}$ as follows.

We set

$$\mathbf{w_h}(\mathbf{x_k}) = \frac{\sum\limits_{s=1}^{N_k} \left|\mathbf{T}_s^{(k)}\right| \cdot \widehat{\mathbf{u}}_h\Big|_{T_s^{(k)}}}{\sum\limits_{s=1}^{N_k} \left|\mathbf{T}_s^{(k)}\right|} - \mathbf{u_0}(\mathbf{x_k}) \tag{9.34}$$

for any internal node $\mathbf{x_k}$ and when $\mathbf{x_k} \in \partial_2\Omega$. Here $\mathbf{T}_s^{(k)}$, $s = \overline{1, N_k}$, 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 $\mathbf{x_k} \in \partial_1\Omega$, we simply set $\mathbf{w_h}(\mathbf{x_k}) = \mathbf{0}$.

Thus, using the nodal values of $\mathbf{w_h}$ and the piecewise-linear continuous finite element approximation on the mesh $\mathcal{T}_h$ we define the function

$$\mathbf{w_h} + \mathbf{u_0} = \mathcal{P}_h \widehat{\mathbf{u}}_h \in \mathbf{V_0} + \mathbf{u_0}\,.$$

Hence, from (9.33) one obtains

$$\bar{\mathbf{c}}_1\|\mathbf{p}-\widehat{\mathbf{p}}_h\| \leq\| \mathbf{p}-\widehat{\mathbf{p}}_h \|_*\leq$$
$$\sqrt{2}\mathbf{D}^{1/2}(\nabla(\mathcal{P}_h\widehat{\mathbf{u}}_h),\widehat{\mathbf{p}}_h)+2\mathbf{C}\left(\|\mathbf{f}-[\mathbf{f}]\|^2+\|\mathbf{F}-[\mathbf{F}]\|^2_{\partial_2\Omega}\right)^{1/2}, \tag{9.35}$$

which, together with the obvious relation

$$\|\mathbf{div}(\widehat{\mathbf{p}}-\widehat{\mathbf{p}}_h)\|=\|-\mathbf{f}-\mathbf{div}\widehat{\mathbf{p}}_h\|=\|\mathbf{f}-[\mathbf{f}]\|$$

leads to the upper bound for $\|\widehat{\mathbf{p}}-\widehat{\mathbf{p}}_h\|_{\mathbf{div}}$:

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* $\mathbf{RT^0}$*. Then, the following estimate holds true:*

$$\|\widehat{\mathbf{p}} - \widehat{\mathbf{p}}_{\mathbf{h}}\|_{\mathbf{div}} \leq \\ \interleave \mathbf{A}\nabla(\mathcal{P}_{\mathbf{h}}\widehat{\mathbf{u}}_{\mathbf{h}}) - \widehat{\mathbf{p}}_{\mathbf{h}} \interleave_{*} + (\mathbf{2C}+\mathbf{1})\|\mathbf{f} - [\mathbf{f}]\| + \mathbf{2C}\|\mathbf{F} - [\mathbf{F}]\|_{\partial_{\mathbf{2}}\boldsymbol{\Omega}}\,, \tag{9.36}$$

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 (9.36), being computed elementwise, can serve as local error indicators.

A sharper estimate can be obtained by the minimization of the Majorant with respect to **v**. Here, we can restrict ourselves to certain subspace V_h, i.e.,

$$\|\widehat{\mathbf{p}} - \widehat{\mathbf{p}}_{\mathbf{h}}\|_{\mathbf{div}} \leq$$
$$\inf_{\mathbf{v_h} \in \mathbf{V_h}} \left\| \mathbf{A}\nabla(\mathbf{v_h}) - \widehat{\mathbf{p}}_{\mathbf{h}} \right\|_{*} + (\mathbf{2C} + \mathbf{1})\|\mathbf{f} - [\mathbf{f}]\| + \mathbf{2C}\|\mathbf{F} - [\mathbf{F}]\|_{\partial_2 \Omega} . \tag{9.37}$$

By (9.28) we can also the squared norm of the error of the averaged solution $\mathcal{P}_h\widehat{\mathbf{u}}_h$ using the computed flux approximation $\widehat{\mathbf{p}}_h$:

$$\|\!| \nabla(\mathbf{u} - \mathcal{P}_h\widehat{\mathbf{u}}_h) \|\!|^2 \leq 2(1+\beta)\mathbf{D}(\nabla(\mathcal{P}_h\widehat{\mathbf{u}}_h), \widehat{\mathbf{p}}_h) + \left(1 + \frac{1}{\beta}\right) \mathbf{C}^2(\|\mathbf{f} - [\mathbf{f}]\|^2 + \|\mathbf{F} - [\mathbf{F}]\|^2_{\partial_2\Omega}), \quad (9.38)$$

where $\beta > 0$ is an arbitrary number that can be used to minimize the right-hand side of (9.38) 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 (9.38) 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 (9.28) to immediately yield the error estimate for the primal variable (which is the most important in the primal mixed method), while the approximation $\mathbf{u_h}$ can be used in (9.36) to bring the error estimate for the dual variable (which is the most important in the dual mixed method).

Lecture 10.
A POSTERIORI ERROR ESTIMATES FOR ITERATION METHODS

Lecture plan

- Banach fixed point theorem;
- Two-sided error estimates by A. Ostrovski;
- Advanced two-sided estimates;
- Applications to matrix equations;
- Positivity methods and a posteriori error bounds.
- Applications to integral equations;
- Applications to ordinary differential equations.

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

$$\mathbf{x}_\odot = \mathfrak{T}\mathbf{x}_\odot\,. \tag{10.1}$$

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

$$\mathbf{x_i} = \mathfrak{T}\mathbf{x_{i-1}} \qquad \mathbf{i} = \mathbf{1}, \mathbf{2}, \ldots\,. \tag{10.2}$$

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 $\mathbf{e_i} = \mathbf{d}(\mathbf{x_i}, \mathbf{x}_\odot)$.

These problems possess solutions provided, that $\mathfrak{T}$ is subject to the following additional condition.

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

$$\mathbf{d}(\mathfrak{T}\mathbf{x}, \mathfrak{T}\mathbf{y}) \le \mathbf{q}\,\mathbf{d}(\mathbf{x}, \mathbf{y}) \tag{10.3}$$

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

Stefan Banach

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 (10.2) converges to this point.

Proof. It is easy to see that

$$\mathbf{d}(\mathbf{x}_{\mathbf{i}+1},\mathbf{x}_{\mathbf{i}}) = \mathbf{d}(\mathfrak{T}\mathbf{x}_{\mathbf{i}},\mathfrak{T}\mathbf{x}_{\mathbf{i}-1}) \leq \mathbf{q}\mathbf{d}(\mathbf{x}_{\mathbf{i}},\mathbf{x}_{\mathbf{i}-1}) \leq \ldots \leq \mathbf{q}^{\mathbf{i}}\mathbf{d}(\mathbf{x}_1,\mathbf{x}_0).$$

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

$$\begin{aligned} &\mathbf{d}(\mathbf{x}_{\mathbf{i}+\mathbf{m}},\mathbf{x}_{\mathbf{i}}) \leq \\ &\leq \mathbf{d}(\mathbf{x}_{\mathbf{i}+\mathbf{m}},\mathbf{x}_{\mathbf{i}+\mathbf{m}-1}) + \mathbf{d}(\mathbf{x}_{\mathbf{i}+\mathbf{m}-1},\mathbf{x}_{\mathbf{i}+\mathbf{m}-2}) + \ldots + \mathbf{d}(\mathbf{x}_{\mathbf{i}+1},\mathbf{x}_{\mathbf{i}}) \leq \\ &\leq \mathbf{q}^{\mathbf{i}}(\mathbf{q}^{\mathbf{m}-1} + \mathbf{q}^{\mathbf{m}-2} + \ldots + 1)\mathbf{d}(\mathbf{x}_1,\mathbf{x}_0)\,. \end{aligned} \tag{10.4}$$

Since

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

(10.4) implies the estimate

$$d(x_{i+m}, x_i) \;\leq\; \frac{q^i}{1-q} d(x_1, x_0). \tag{10.5}$$

Let $\mathbf{i} \to \infty$, then the right-hand side of (10.5) tends to zero, so that $\{x_i\}$ is a Cauchy sequence. It has a limit in $\mathbf{y} \in \mathbf{X}$.

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

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

so that $\mathbf{d}(\mathfrak{T}\mathbf{x_i}, \mathfrak{T}\mathbf{y}) \longrightarrow \mathbf{0}$ and $\mathfrak{T}\mathbf{x_i} \longrightarrow \mathfrak{T}\mathbf{y}$. Pass to the limit in (10.2) as $\mathbf{i} \longrightarrow +\infty$. We observe that

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

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.

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

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

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

This estimate gives a certain presentation on that how the error decreases. However, as we will see later, 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.

Theorem (A. Ostrowski)

Let $\{\mathbf{x_j}\}_{\mathbf{j=0}}^{\infty}$ *be a sequence obtained by the iteration process (10.2) with a mapping* $\mathfrak{T}$ *satisfying the condition (10.3). Then, for any* $\mathbf{x_j}$, $\mathbf{j} > \mathbf{1}$, *the following estimate holds:*

$$\mathbf{M}_{\ominus}^{\mathbf{j}} := \frac{\mathbf{1}}{\mathbf{1+q}}\mathbf{d}(\mathbf{x_{j+1}}, \mathbf{x_j}) \leq \mathbf{e_j} \leq \mathbf{M}_{\oplus}^{\mathbf{j}} := \frac{\mathbf{q}}{\mathbf{1-q}}\mathbf{d}(\mathbf{x_j}, \mathbf{x_{j-1}}) \quad (10.6)$$

A. Ostrowski

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

Proof. The upper estimate in (10.6) follows from (10.5). Indeed, put $\mathbf{i} = \mathbf{1}$ in this relation. We have

$$\mathbf{d}(\mathbf{x}_{\mathbf{1}+\mathbf{m}}, \mathbf{x}_{\mathbf{1}}) \leq \frac{\mathbf{q}}{\mathbf{1} - \mathbf{q}} \mathbf{d}(\mathbf{x}_{\mathbf{1}}, \mathbf{x}_{\mathbf{0}})\,.$$

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

$$\mathbf{d}(\mathbf{x}_{\odot}, \mathbf{x}_{\mathbf{1}}) \leq \frac{\mathbf{q}}{\mathbf{1} - \mathbf{q}} \mathbf{d}(\mathbf{x}_{\mathbf{1}}, \mathbf{x}_{\mathbf{0}})\,.$$

We may view $\mathbf{x}_{\mathbf{j}-\mathbf{1}}$ as the starting point of the sequence. Then, in the above relation $\mathbf{x}_{\mathbf{0}} = \mathbf{x}_{\mathbf{j}-\mathbf{1}}$ and $\mathbf{x}_{\mathbf{1}} = \mathbf{x}_{\mathbf{j}}$ and we arrive at the following upper bound of the error:

$$\mathbf{d}(\mathbf{x}_{\odot}, \mathbf{x}_{\mathbf{j}}) \leq \frac{\mathbf{q}}{\mathbf{1} - \mathbf{q}} \mathbf{d}(\mathbf{x}_{\mathbf{j}}, \mathbf{x}_{\mathbf{j}-\mathbf{1}})\,.$$

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 (10.6) 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 (10.1). 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:

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

Corollaries

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

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

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 $\xi_\odot$.
Let us show that $\xi_\odot$ is a fixed point of $\mathbf{T}$, we note that

$$\begin{aligned}\mathbf{T}\xi_\odot = \mathbf{T}(\mathfrak{T}\xi_\odot) = \mathbf{T}\,\mathfrak{T}^2\xi_\odot = \ldots \\ = \mathbf{T}\,\mathfrak{T}^{\mathbf{i}}\xi_\odot = \mathbf{T}^{(1+\mathbf{in})}\xi_\odot = \mathbf{T}^{\mathbf{in}}\mathbf{T}\xi_\odot.\end{aligned} \quad (10.7)$$

Denote $\mathbf{x_0} = \mathbf{T}\xi_\odot$. By (10.7) we conclude that for any $\mathbf{i}$

$$\mathbf{T}\xi_\odot = \mathfrak{T}^{\mathbf{i}}\mathbf{x_0}. \quad (10.8)$$

Passing to the limit on the right-hand side in (10.8), we arrive at the relation $\mathbf{T}\xi_\odot = \xi_\odot$, which means that $\xi_\odot$ 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 $\mathbf{T}$.

Advanced two–sided a posteriori estimates

We can derive more accurate bounds of errors if we use **more terms of the sequence** $\{x_j\}$.
Indeed,

$$\begin{aligned} d(x_j, x_\odot) \leq \quad & d(x_j, x_{j+1}) + d(x_{j+1}, x_\odot) \leq \\ & \leq d(x_j, x_{j+1}) + \frac{q}{1-q} d(x_j, x_{j+1}), \end{aligned}$$

and we obtain another upper bound

$$d(x_j, x_\odot) \leq \frac{1}{1-q} d(x_j, x_{j+1}). \tag{10.9}$$

It estimates the error on j-th step by x_j and x_{j+1}.

Which bound is sharper: (10.9) or $\mathbf{M}^{\mathbf{j}}_{\oplus}$?

Since

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

we observe that this bound is sharper than $\mathbf{M}^{\mathbf{j}}_{\oplus}$.
Obviously, (10.9) can also be applied to *any subsequence of* $\{\mathbf{x_j}\}$. For example, we can take $\{\mathbf{x}_{\ell \mathbf{s}}\}$, $\mathbf{s}=\mathbf{0},\mathbf{1},\mathbf{2}...$ with some fixed ℓ. In this case, we obtain various upper bounds of $\mathbf{d}(\mathbf{x_j},\mathbf{x}_{\odot})$ computed on the basis of some terms of the sequence $\{\mathbf{x_j}\}$:

$$\mathbf{d}(\mathbf{x_j},\mathbf{x}_{\odot}) \leq \mathbf{M}^{\mathbf{j},\ell}_{\oplus} := \frac{\mathbf{1}}{\mathbf{1}-\mathbf{q}^{\ell}}\mathbf{d}(\mathbf{x_j},\mathbf{x}_{\mathbf{j}+\ell}).$$

Note that the right-hand side of this estimate tends to $\mathbf{d}(\mathbf{x_j},\mathbf{x}_{\odot})$ as $\ell \to +\infty$. Thus, for a sufficiently large ℓ the bound will be accurate even if $\mathbf{q}$ is close to 1.

The lower estimates can be improved by similar arguments. We have the estimate

$$\mathbf{d}(\mathbf{x_j}, \mathbf{x}_\odot) \geq \mathbf{M}_\ominus^{\mathbf{j},\ell} := \frac{\mathbf{1}}{\mathbf{1}+\mathbf{q}^\ell}\mathbf{d}(\mathbf{x_j}, \mathbf{x}_{\mathbf{j}+\ell})$$

whose right-hand side also tends to the exact value of the error as $\ell \to +\infty$.

Let $\mathbf{L}$ be a given number that indicates the number of successive elements used in the evaluation of error bounds for $\mathbf{x_j}$. Compute the quantities

$$\bar{\mathbf{M}}_\ominus^{\mathbf{j},\mathbf{L}} := \sup_{\ell=\mathbf{1},\mathbf{2},\ldots\mathbf{L}} \left\{ \frac{\mathbf{1}}{\mathbf{1}+\mathbf{q}^\ell}\,\mathbf{d}(\mathbf{x_j}, \mathbf{x}_{\mathbf{j}+\ell}) \right\}, \tag{10.10}$$

$$\bar{\mathbf{M}}_\oplus^{\mathbf{j},\mathbf{L}} := \inf_{\ell=\mathbf{1},\mathbf{2},\ldots\mathbf{L}} \left\{ \frac{\mathbf{1}}{\mathbf{1}-\mathbf{q}^\ell}\,\mathbf{d}(\mathbf{x_j}, \mathbf{x}_{\mathbf{j}+\ell}) \right\}. \tag{10.11}$$

These upper and lower bounds are the sharper, the greater is $\mathbf{L}$.

Another sequence of upper bounds follows from the relation

$$\mathbf{d}(\mathbf{x_j}, \mathbf{x_\odot}) \leq \mathbf{d}(\mathbf{x_j}, \mathbf{x_{j+1}}) + \mathbf{d}(\mathbf{x_{j+1}}, \mathbf{x_\odot}) \leq$$
$$\leq \mathcal{M}_\oplus^{\mathbf{j},2}(\mathbf{x_j}, \mathbf{x_{j+1}}, \mathbf{x_{j+2}}) := \mathbf{d}(\mathbf{x_j}, \mathbf{x_{j+1}}) + \frac{\mathbf{1}}{\mathbf{1}-\mathbf{q}}\mathbf{d}(\mathbf{x_{j+1}}, \mathbf{x_{j+2}}). \quad (10.12)$$

Note that

$$\mathcal{M}_\oplus^{\mathbf{j},2} \leq \mathbf{d}(\mathbf{x_j}, \mathbf{x_{j+1}}) + \frac{\mathbf{q}}{\mathbf{1}-\mathbf{q}}\mathbf{d}(\mathbf{x_j}, \mathbf{x_{j+1}}) =$$
$$= \frac{\mathbf{1}}{\mathbf{1}-\mathbf{q}}\mathbf{d}(\mathbf{x_j}, \mathbf{x_{j+1}}) := \mathbf{M}_\oplus^{\mathbf{j},\mathbf{1}}.$$

Similarly, we can obtain lower bounds of the error computed by $\mathbf{x_j}$, $\mathbf{x_{j+1}}$, and $\mathbf{x_{j+2}}$.

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

$$\mathbf{x_j} = \mathcal{L}\,\mathbf{x_{j-1}} + \mathbf{b}. \tag{10.13}$$

Let $\mathbf{x}_{\odot}$ be a fixed point of (10.13) 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

$$\bar{\mathbf{x}}_\mathbf{j} = \mathcal{L}\mathbf{x_{j-1}} + \mathbf{b} - \mathbf{x}_\odot = \mathcal{L}(\mathbf{x_{j-1}} - \mathbf{x}_\odot) = \mathcal{L}\bar{\mathbf{x}}_\mathbf{j-1} \tag{10.14}$$

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 $\mathcal{L}$.

Therefore, we have an *a priori* estimate

$$\begin{aligned} \|\mathbf{x_j} - \mathbf{x_\odot}\|_{\mathbf{X}} = \|\bar{\mathbf{x}}_\mathbf{j} - \mathbf{0_X}\|_{\mathbf{X}} \leq \\ \leq \frac{\mathbf{q^j}}{\mathbf{1-q}} \|\bar{\mathbf{x}}_\mathbf{1} - \bar{\mathbf{x}}_\mathbf{0}\|_{\mathbf{X}} = \frac{\mathbf{q^j}}{\mathbf{1-q}} \|\mathbf{R}(\mathbf{x_0})\|_{\mathbf{X}} \end{aligned} \quad (10.15)$$

and the *a posteriori* one

$$\|\mathbf{x_j} - \mathbf{x_\odot}\|_{\mathbf{X}} \leq \frac{\mathbf{q}}{\mathbf{1-q}} \|\mathbf{R}(\mathbf{x_{j-1}})\|_{\mathbf{X}}\,, \quad (10.16)$$

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

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

$$\|x_j - x_\odot\|_X \geq \frac{1}{1+q}\|x_{j+1} - x_j\|_X = \frac{1}{1+q}\|R(x_j)\|_X . \quad (10.17)$$

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

$$\frac{1-q}{q}\|x_j - x_\odot\|_X \leq \|R(x_{j-1})\|_X \leq (1+q)\|x_{j-1} - x_\odot\|_X .$$

Advanced estimates that provide sharper bounds can be easily obtained by applying (10.10) and (10.11).

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 **A** are often represented in the form

$$\mathbf{B}\frac{\mathbf{x_i} - \mathbf{x_{i-1}}}{\tau} + \mathbf{A}\,\mathbf{x_{i-1}} = \mathbf{f}\,. \tag{10.18}$$

In (10.18), the matrix **B** and the parameter τ may be taken in various ways (depending on the properties of **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 (10.18) is reduced to (10.13) if we set

$$\mathcal{L} = \mathbb{I} - \boldsymbol{\tau}\mathbf{B}^{-1}\mathbf{A} \qquad \text{and} \qquad \mathbf{b} = \boldsymbol{\tau}\mathbf{B}^{-1}\mathbf{f}\,, \tag{10.19}$$

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

$$\mathbf{A}\,\mathbf{x}_{\odot} = \mathbf{f} \tag{10.20}$$

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 (10.15)-(10.17), the quantities

$$\begin{aligned}
\mathbf{M}_{\oplus}^{\mathbf{i}} = &\quad \mathbf{q}(\mathbf{1}-\mathbf{q})^{-\mathbf{1}}\,\|\mathbf{R}(\mathbf{x}_{\mathbf{i}-\mathbf{1}})\|\,, && (10.21)\\
\mathbf{M}_{\oplus}^{\mathbf{0i}} = &\quad \mathbf{q}^{\mathbf{i}}(\mathbf{1}-\mathbf{q})^{-\mathbf{1}}\,\|\mathbf{R}(\mathbf{x}_{\mathbf{0}})\|\,, && (10.22)\\
\mathbf{M}_{\ominus}^{\mathbf{i}} = &\quad (\mathbf{1}+\mathbf{q})^{-\mathbf{1}}\,\|\mathbf{R}(\mathbf{x}_{\mathbf{i}})\| && (10.23)
\end{aligned}$$

furnish upper and lower bounds of the error for the vector $\mathbf{x}_{\mathbf{i}}$. The validity of them is demonstrated with an example below.
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.

Remark. If $\mathbf{q}$ 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* $\mathbf{P}^{-1}$ and solving the system

$$\left(\mathbf{P}^{-1}\mathbf{A}\right)\mathbf{x} = \mathbf{P}^{-1}\mathbf{f}$$

with a smaller condition number. Of cause, the best preconditioner is the unknown matrix $\mathbf{A}^{-1}$. Therefore, a preconditioner is often constructed from the parts of $\mathbf{A}$ that are not difficult to invert (e.g., in the simplest case it is taken as the matrix inverse to the diagonal part of $\mathbf{A}$). This iteration technique is well presented in the literature: see, e.g.,

O. Axelsson. *Iterative solution methods*. Cambridge University Press, Cambridge, 1994.

Examples

Consider the problem $\mathbf{Ax} = \mathbf{f}$ for a symmetric matrix $\mathbf{A}$ with coefficients $\mathbf{a_{ij}} = \mathbf{0.8/ij}$ if $\mathbf{i} \neq \mathbf{j}$ and $\mathbf{a_{ii}} = \mathbf{i}$. The system is solved by the method

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

with $\mathbf{B} = \mathbf{A_D}$ and $\mathbf{x_0} = \{\mathbf{0}, \mathbf{0}, ...\mathbf{0}\}$.
In this example $\mathbf{n} = 200$, $\mathbf{q} = 0.662$, and $\boldsymbol{\tau} = 0.760$. The values of the error and the estimates are presented below.

Table:

i	$\mathbf{M}^{\mathbf{i}}_{\ominus}$	$\|\mathbf{e}\|$	$\mathbf{M}^{\mathbf{i}}_{\oplus}$	$\mathbf{M}^{\mathbf{0i}}_{\oplus}$
1	.187145E+03	.412471E+03	.245893E+04	.245893E+04
2	.452820E+02	.104019E+03	.610732E+03	.162904E+04
3	.123433E+02	.311517E+02	.147774E+03	.107924E+04
4	.405504E+01	.116679E+02	.402813E+02	.714995E+03
5	.166633E+01	.517711E+01	.132333E+02	.473684E+03
6	.767379E+00	.244532E+01	.543792E+01	.313815E+03
7	.366283E+00	.117450E+01	.250428E+01	.207902E+03
8	.176340E+00	.566166E+00	.119533E+01	.137735E+03
16	.515722E-03	.165576E-02	.349042E-02	.511127E+01
17	.248671E-03	.798371E-03	.168302E-02	.338621E+01
18	.119903E-03	.384956E-03	.811515E-03	.224336E+01
19	.578146E-04	.185617E-03	.391295E-03	.148623E+01
20	.278769E-04	.895001E-04	.188673E-03	.984624E+00

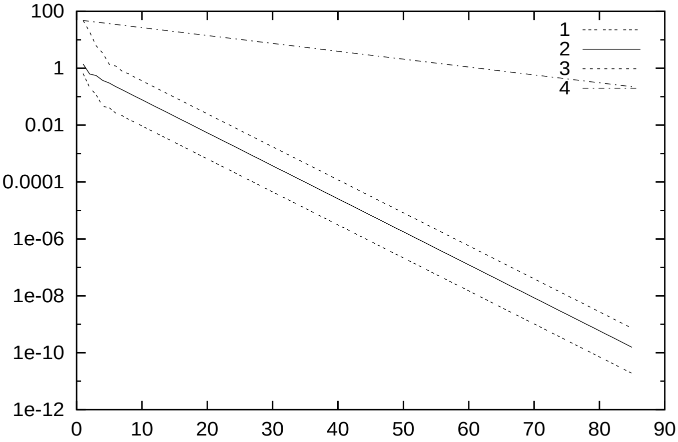


Figure: A priori and a posteriori estimates for an iteration process: $1 - \mathbf{M}^{\mathbf{i}}_{\ominus}$, $2 - \|\mathbf{e}\|$, $3 - \mathbf{M}^{\mathbf{i}}_{\oplus}$, $4 - \mathbf{M}^{\mathbf{0i}}_{\oplus}$.

Positivity methods and a posteriori error bounds.

In some cases, one can obtain two-sided estimates *for each component of a solution*. The respective methods can be viewed as a simplest example of the so–called positivity methods widely used in the analysis of differential equations.

Let $\mathbf{x}_\odot$ be a solution of the system of linear simultaneous equations

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

where

$$\mathbf{A} = \mathbf{A}^\oplus - \mathbf{A}^\ominus$$

and

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

We may *partially order* the space $\mathbb{R}^{\mathbf{n}}$ by saying that $\mathbf{x} \leq \mathbf{y}$ if and only if $\mathbf{x_i} \leq \mathbf{y_i}$ for $i = 1, 2, \ldots n$.
Assume that the vectors $\mathbf{x}_{\mathbf{0}}^{\ominus}$ and $\mathbf{x}_{\mathbf{0}}^{\oplus}$ are ordered such that

$$\mathbf{x}_{\mathbf{0}}^{\ominus} \leq \mathbf{x}_{\odot} \leq \mathbf{x}_{\mathbf{0}}^{\oplus}.$$

The vectors $\mathbf{x}_{\mathbf{0}}^{\ominus}$ and $\mathbf{x}_{\mathbf{0}}^{\oplus}$ are considered as the *initial guesses* for the bounds of the solution components.
Compute $\mathbf{x}_{\mathbf{1}}^{\ominus}$ and $\mathbf{x}_{\mathbf{1}}^{\oplus}$ by the relations

$$\mathbf{x}_{\mathbf{1}}^{\ominus} = \mathbf{A}^{\oplus}\mathbf{x}_{\mathbf{0}}^{\ominus} - \mathbf{A}^{\ominus}\mathbf{x}_{\mathbf{0}}^{\oplus} + \mathbf{f},$$
$$\mathbf{x}_{\mathbf{1}}^{\oplus} = \mathbf{A}^{\oplus}\mathbf{x}_{\mathbf{0}}^{\oplus} - \mathbf{A}^{\ominus}\mathbf{x}_{\mathbf{0}}^{\ominus} + \mathbf{f}.$$

It is easy to see that

$$\mathbf{x}_1^{\ominus} - \mathbf{x}_{\odot} = \mathbf{A}^{\oplus}(\mathbf{x}_0^{\ominus} - \mathbf{x}_{\odot}) - \mathbf{A}^{\ominus}(\mathbf{x}_0^{\oplus} - \mathbf{x}_{\odot}) \leq \mathbf{0},$$
$$\mathbf{x}_1^{\oplus} - \mathbf{x}_{\odot} = \mathbf{A}^{\oplus}(\mathbf{x}_0^{\oplus} - \mathbf{x}_{\odot}) - \mathbf{A}^{\ominus}(\mathbf{x}_0^{\ominus} - \mathbf{x}_{\odot}) \geq \mathbf{0}.$$

Hence,

$$\mathbf{x}_1^{\ominus} \leq \mathbf{x}_{\odot} \leq \mathbf{x}_1^{\oplus}.$$

and we observe that $\mathbf{x}_1^{\ominus}$ and $\mathbf{x}_1^{\oplus}$ also give *componentwise bounds* for the exact solution.

Quite similarly, we observe that the subsequent elements of the iteration process

$$\mathbf{x}_{\mathbf{k+1}}^{\ominus} = \mathbf{A}^{\oplus}\mathbf{x}_{\mathbf{k}}^{\ominus} - \mathbf{A}^{\ominus}\mathbf{x}_{\mathbf{k}}^{\oplus} + \mathbf{f}$$
$$\mathbf{x}_{\mathbf{k+1}}^{\oplus} = \mathbf{A}^{\oplus}\mathbf{x}_{\mathbf{k}}^{\oplus} - \mathbf{A}^{\ominus}\mathbf{x}_{\mathbf{k}}^{\ominus} + \mathbf{f},$$

possess the same properties. Therefore, for the ith component we find the following two-sided bounds:

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

Similar methods can be applied to *functional equations*, provided that the operator $\mathbf{A}$ is presented as the sum of

$$\mathbf{A}^{\oplus} \quad \text{and} \quad (-\mathbf{A}^{\ominus})$$

which are certain monotone operators defined on a partially ordered space:
see, e.g.,
L. Collatz. *Funktionanalysis und numerische mathematik*. Springer-Verlag, Berlin, 1964.

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 $\mathbf{x}_{\odot}(\mathbf{t}) \in \mathbf{C}[\mathbf{a}, \mathbf{b}]$ such that

$$\mathbf{x}_{\odot}(\mathbf{t}) = \lambda \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K}(\mathbf{t}, \mathbf{s})\, \mathbf{x}_{\odot}(\mathbf{s})\, \mathbf{ds} + \mathbf{f}(\mathbf{t}), \tag{10.24}$$

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

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

and

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

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

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

$$\mathbf{y}(\mathbf{t}) := \mathfrak{T}\mathbf{x}(\mathbf{t}) := \lambda \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K}(\mathbf{t}, \mathbf{x})\mathbf{x}(\mathbf{s})\, \mathbf{ds} + \mathbf{f}(\mathbf{t}) \qquad (10.25)$$

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

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

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

Therefore, for any given ε one can find a small number $\boldsymbol{\delta}$ such that

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

and

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

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

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

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

$\mathfrak{T} : \mathbf{C}[\mathbf{a},\mathbf{b}] \to \mathbf{C}[\mathbf{a},\mathbf{b}]$ 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

$$\mathbf{q} = |\lambda|\mathbf{M}(\mathbf{b}-\mathbf{a}), \tag{10.26}$$

provided that

$$|\lambda| < \frac{1}{\mathbf{M}(\mathbf{b}-\mathbf{a})}. \tag{10.27}$$

Numerical procedure

An approximate solution of (10.24) can be found by the iteration method

$$\mathbf{x_{i+1}(t)} = \lambda \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K(t,s)x_i(s)\,ds + f(t)}. \qquad (10.28)$$

If (10.27) 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

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

$$\leq \max_{\mathbf{a \leq t \leq b}} |\mathbf{x_i(t) - x_\odot(t)}| \leq \frac{\mathbf{q}}{\mathbf{1-q}} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{K(t,s)(x_i(s) - x_{i-1}(s))\,ds}. \qquad (10.29)$$

Applications to Volterra type equations

Consider the fixed point problem

$$\mathbf{x}_{\odot}(\mathbf{t}) = \lambda \int_{\mathbf{a}}^{\mathbf{t}} \mathbf{K}(\mathbf{t},\mathbf{s})\, \mathbf{x}_{\odot}(\mathbf{s})\, \mathbf{ds} + \mathbf{f}(\mathbf{t}), \qquad (10.30)$$

where

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

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

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

By the same arguments we find that

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

Thus, the operator $\mathfrak{T} := \mathbf{T^n}$ 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

$$\frac{\mathbf{du}}{\mathbf{dt}} = \varphi(\mathbf{t}, \mathbf{u}(\mathbf{t})), \quad \mathbf{u}(\mathbf{t_0}) = \mathbf{a}, \tag{10.31}$$

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

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

and

$$|\varphi(\mathbf{t}, \mathbf{p_1}) - \varphi(\mathbf{t}, \mathbf{p_2})| \leq \mathbf{L}|\mathbf{p_1} - \mathbf{p_2}|, \quad \forall(\mathbf{t}, \mathbf{p}) \in \mathbf{Q}. \tag{10.32}$$

Problem (10.31) can be reduced to the integral equation

$$\mathbf{u}(\mathbf{t}) = \int_{\mathbf{t_0}}^{\mathbf{t}} \varphi(\mathbf{s}, \mathbf{u}(\mathbf{s}))\, \mathbf{ds} + \mathbf{a} \tag{10.33}$$

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

$$\mathbf{u_j}(\mathbf{t}) = \int_{\mathbf{t_0}}^{\mathbf{t}} \varphi(\mathbf{s}, \mathbf{u_{j-1}}(\mathbf{s}))\, \mathbf{ds} + \mathbf{a}. \tag{10.34}$$

To justify this procedure, we must verify that the operator

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

is $\mathbf{q}$-contractive with respect to the norm

$$\|\mathbf{u}\| := \max_{\mathbf{t}\in[\mathbf{t_0},\mathbf{t_1}]} |\mathbf{u}(\mathbf{t})|. \tag{10.35}$$

We have

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

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

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

We see that if

$$t_1 < t_0 + L^{-1}, \tag{10.36}$$

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

$$q := L(t_1 - t_0) < 1.$$

Therefore, if the interval $[\mathbf{t_0}, \mathbf{t_1}]$ is small enough (i.e., it satisfies the condition (10.36), then the existence and uniqueness of a continuous solution $\mathbf{u}(\mathbf{t})$ 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.

Lecture 11.
A POSTERIORI ESTIMATES FOR VARIATIONAL INEQUALITIES

Lecture plan

- Variational inequalities. Background;
- Deviation estimates for variational inequalities;
- Obstacle problem;
- Functional type a posteriori estimates for problems with two obstacles;
- Examples;
- Elasto-plastic torsion problem;

Variational inequalities

Variational inequalities provide a mathematical description of a vide spectrum of nonlinear boundary–value problems that arise in various applications (see, e.g., **G. Duvant and J.-L. Lions.** *Les inequations en mecanique et en physique*, Dunod, Paris, 1972.)

First we establish the relationship between variational inequalities and certain variational problems. Consider the functional

$$\mathbf{J}(\mathbf{v}) = \mathbf{J_0}(\mathbf{v}) + \mathbf{j}(\mathbf{v}),$$

where $\mathbf{J_0} : \mathbf{V} \longrightarrow \mathbb{R}$ is a convex, continuous, and Gateaux-differentiable functional and $\mathbf{j}(\mathbf{v}) : \mathbf{V} \longrightarrow \mathbb{R}$ is a convex and continuous functional.

Let $\mathbf{K}$ be a convex closed subset of a reflexive Banach space $\mathbf{V}$. Consider the following problem: find $\mathbf{u} \in \mathbf{K}$ such that

$$\mathbf{J}(\mathbf{u}) = \inf_{\mathbf{v}\in\mathbf{K}} \mathbf{J}(\mathbf{v}), \quad \mathbf{J}(\mathbf{v}) = \mathbf{J_0}(\mathbf{v}) + \mathbf{j}(\mathbf{v}). \tag{11.1}$$

Hereafter, we assume that $\mathbf{J}$ is coercive on $\mathbf{V}$, so that the above problem has a solution $\mathbf{u}$.
Moreover, the minimizer satisfies the relation

$$\left(\mathbf{J}_0'(\mathbf{u}), \mathbf{u} - \mathbf{v}\right) + \mathbf{j}(\mathbf{u}) - \mathbf{j}(\mathbf{v}) \leq \mathbf{0} \qquad \forall \mathbf{v} \in \mathbf{K} \tag{11.2}$$

Theorem (1)

Relations (11.1) and (11.2) are equivalent.

Proof

1. Let(11.1) holds, i.e.

$$\mathbf{J_0}(\mathbf{v}) + \mathbf{j}(\mathbf{v}) \geq \mathbf{J_0}(\mathbf{u}) + \mathbf{j}(\mathbf{u}) \qquad \forall \mathbf{v} \in \mathbf{K}.$$

Take $\mathbf{v} = \mathbf{u} + \lambda(\mathbf{w} - \mathbf{u})$, $\mathbf{w} \in \mathbf{K}$, $\lambda \in [0, 1]$.
Then

$$\mathbf{J_0}\left(\mathbf{u} + \lambda(\mathbf{w} - \mathbf{u})\right) - \mathbf{J_0}(\mathbf{u}) + \mathbf{j}(\mathbf{v}) - \mathbf{j}(\mathbf{u}) \geq \mathbf{0} \qquad \forall \mathbf{u} \in \mathbf{K}.$$

By the convexitity of $\mathbf{j}$ we have

$$\begin{aligned} \mathbf{j}(\mathbf{v}) = \mathbf{j}\left(\mathbf{u} + \lambda(\mathbf{w} - \mathbf{u})\right) &= \mathbf{j}\left(\lambda\mathbf{w} + (\mathbf{1} - \lambda)\mathbf{u}\right) \\ &\leq \lambda\mathbf{j}(\mathbf{w}) + (\mathbf{1} - \lambda)\mathbf{j}(\mathbf{u}). \end{aligned}$$

Thus, for any $w \in \mathbf{K}$ we have

$$\begin{aligned} \mathbf{J_0}\left(\mathbf{u}+\lambda(\mathbf{w}-\mathbf{u})\right)-\mathbf{J_0}(\mathbf{u})+\lambda\mathbf{j}(\mathbf{w}) \\ +(\mathbf{1}-\lambda)\mathbf{j}(\mathbf{u})-\mathbf{j}(\mathbf{u})\geq \mathbf{0} \qquad \forall \mathbf{w}\in\mathbf{K}, \\ \frac{\mathbf{1}}{\lambda}\left(\mathbf{J_0}\left(\mathbf{u}+\lambda(\mathbf{w}-\mathbf{u})\right)-\mathbf{J_0}(\mathbf{u})\right) \\ +\mathbf{j}(\mathbf{w})-\mathbf{j}(\mathbf{u})\geq \mathbf{0} \qquad \forall \mathbf{u}\in\mathbf{K}. \end{aligned}$$

Passing to the limit as $\lambda \to 0$ we obtain

$$\left(\mathbf{J}_0'(\mathbf{u}),\mathbf{w}-\mathbf{u}\right)+\mathbf{j}(\mathbf{w})-\mathbf{j}(\mathbf{u})\geq \mathbf{0} \qquad \forall \mathbf{w}\in\mathbf{K}.$$

2. Assume now that (11.2) holds For a convex functional $\mathbf{J_0}$ we have the relation

$$\mathbf{J_0(v)} \geq \mathbf{J_0(u)} + \left(\mathbf{J_0'(u)}, \mathbf{v} - \mathbf{u}\right).$$

Since

$$\left(\mathbf{J_0'(u)}, \mathbf{u} - \mathbf{v}\right) + \mathbf{j(u)} - \mathbf{j(v)} \leq \mathbf{0} \qquad \forall \mathbf{v} \in \mathbf{K}$$

and

$$(\mathbf{J_0'(u)}, \mathbf{u} - \mathbf{v}) \geq \mathbf{J_0(u)} - \mathbf{J_0(v)}$$

we find that

$$-\mathbf{J_0(v)} + \mathbf{J_0(u)} + \mathbf{j(u)} - \mathbf{j(v)} \leq \mathbf{0} \qquad \forall \mathbf{v} \in \mathbf{K},$$

what means that

$$\mathbf{J(u)} \leq \mathbf{J(v)} \qquad \forall \mathbf{v} \in \mathbf{K}.$$

Variational inequalities can be regarded as Euler's equations to certain convex variational problems with nondifferentiable functionals defined on convex subsets. If the nondifferentiable part of such a functional vanishes and the set coincide with the whole space, then the respective variational inequality converts to a variational equality (integral identity). However, in many practically interesting problems it is impossible to define a mininimizer throughout an integral identity. This fact stimulated the development of the theory of variational inequalities and their numerical analysis (see, e.g.,

R. Glowinski. *Numerical methods for nonlinear variational problems.* Springer-Verlag, New-York, 1982.

Obstacle problem. Introduction

Let $\mathbf{\Omega}$ be a bounded domain in $\mathbb{R}^{\mathbf{n}}$ $(\mathbf{n} = \mathbf{1}, \mathbf{2})$ with L–continuous boundary $\partial\mathbf{\Omega}$ and

$$\mathbf{f} \in \mathbf{L}_2(\mathbf{\Omega}),$$
$$\varphi \in \mathbf{H}^2(\mathbf{\Omega}), \qquad \varphi(\mathbf{x}) \leq \mathbf{0} \text{ on } \partial\mathbf{\Omega}.$$

"Admissible" functions belong to the set

$$\mathbf{K}_\varphi := \{\mathbf{v} \in \mathbf{V} \mid \mathbf{v}(\mathbf{x}) \geq \varphi(\mathbf{x}) \text{ a.e. in } \mathbf{\Omega}\},$$

where

$$\mathbf{V} := \{\mathbf{v} \in \mathbf{H}^1(\mathbf{\Omega}) \mid \mathbf{v} = \mathbf{0} \text{ on } \partial\mathbf{\Omega}\}.$$

Let

$$\mathbf{a}(\mathbf{u},\mathbf{v}) := \int_\Omega \nabla \mathbf{u} \cdot \nabla \mathbf{v}\,\mathbf{dx},$$
$$(\mathbf{u},\mathbf{v}) := \int_\Omega \mathbf{uv}\,\mathbf{dx}.$$

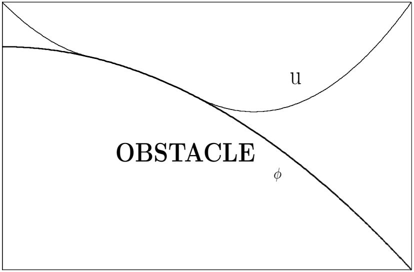
Then, the problem has a variational form

Problem $\mathcal{P}$. Find $\mathbf{u} \in \mathbf{K}_\varphi$ such that

$$\mathbf{J}(\mathbf{u}) = \inf_{\mathbf{v} \in \mathbf{K}_\varphi} \mathbf{J}(\mathbf{v}),$$
$$\mathbf{J}(\mathbf{v}) = \frac{1}{2}\mathbf{a}(\mathbf{v},\mathbf{v}) - (\mathbf{f},\mathbf{v})$$

Physical interpretation

This problem can be interpreted as the one for an elastic membrane deformed at the neighborhood of an obstacle $\varphi(x)$.



Existence of a minimizer

Theorem (Lions – Stampacchia)

Under the above assumptions Problem $\mathcal{P}$ possesses a unique solution $\mathbf{u}$.

Problem $\mathcal{P}$ is, in fact, a **free boundary problem**:

$$\overline{\boldsymbol{\Omega}} = \boldsymbol{\Omega}_\varphi \cup \overline{\boldsymbol{\Omega}_0} \quad \text{coincidenceset}$$

where

$$\boldsymbol{\Omega}_\varphi := \{\mathbf{x} \in \boldsymbol{\Omega} \mid \mathbf{u}(\mathbf{x}) = \varphi(\mathbf{x})\}$$

and

$$\boldsymbol{\Omega}_0 := \{\mathbf{x} \in \boldsymbol{\Omega} \mid \mathbf{u}(\mathbf{x}) > \varphi(\mathbf{x})\}$$

Minimizer $\mathbf{u}$ satisfies the variational inequality

$$\mathbf{a}(\mathbf{u}, \mathbf{v} - \mathbf{u}) \geq (\mathbf{f}, \mathbf{v} - \mathbf{u}) \quad \forall \mathbf{v} \in \mathbf{K}_\varphi.$$

If $\mathbf{u}$ is sufficiently regular, then directly from the variational inequality we derive the following relations that must hold for the solution:

$$\begin{aligned} \boldsymbol{\Delta}\mathbf{u} + \mathbf{f} &= \mathbf{0} && \text{on} \quad \boldsymbol{\Omega}_{\mathbf{0}}, \\ \boldsymbol{\Delta}\mathbf{u} + \mathbf{f} \le \mathbf{0} \quad \mathbf{u} &\ge \varphi && \text{a.e. in } \boldsymbol{\Omega}, \\ (\boldsymbol{\Delta}\mathbf{u} + \mathbf{f})(\mathbf{u} - \varphi) &= \mathbf{0} && \text{a.e. in } \boldsymbol{\Omega}, \end{aligned}$$

Regularity estimates for obstacle problems

In the papers by **H. Brezis, D. Kinderlehrer, H. Lewy, G. Stampacchia** and others, it was shown that

If $\mathbf{f} \in \mathbf{L_2}$ and $\varphi \in \mathbf{H^2}(\boldsymbol{\Omega})$ then $\mathbf{u} \in \mathbf{H^2}(\boldsymbol{\Omega})$.

Moreover, if $\mathbf{f} \in \mathbf{C^1}(\boldsymbol{\Omega})$, $\boldsymbol{\Omega}$ is a bounded domain with smooth boundary, and $\varphi \in \mathbf{C^2}$ than the respective solution possesses second derivatives bounded in $\mathbf{L}_\infty$.

Coincidence set

If $\boldsymbol{\Omega} \subset \mathbf{R}^2$ is strictly convex with smooth boundary and if $\varphi \in \mathbf{C}^2$ is strictly concave, then the coincidence set is connected and its boundary is smooth and homeomorphic to the unit circle.

In general, for any $\boldsymbol{\Omega}$ one can point out such an obstacle φ that $\boldsymbol{\Omega}_\varphi$ has any number of disjoint subsets.

Summary

- Problem $\mathcal{P}$ is related to a variational inequality.
- The coincidence set Ω_φ is unknown a priori, so that a solution has a free boundary.
- Solutions of Problem $\mathcal{P}$ have a bounded regularity even for smooth external data (in the best case scenario the second derivatives are summable, but the third ones are only distributions).

A priori convergence estimates

A priori convergence estimates for problems with obstacles were derived in:

R. S. Falk, *Math. Comp.* 28 (1974),

U. Mosco and G. Strang, *Bull. AMS* 80 (1974),

F. Brezzi, W.W. Hager, P.A. Raviart, *Numer.Math.*28(1997)

It was shown that for regular FE approximations of $\mathbf{u} \in \mathbf{H^2}$:

$$\|\nabla(\mathbf{u} - \mathbf{u_h})\|_{\Omega} \leq \mathbf{C}(\mathbf{u}, \mathbf{f}, \varphi)\, \mathbf{h}$$

A posteriori error estimates for FEM

Two methods usually applied for linear PDEs, namely *residual method* and *gradient averaging methods* are difficult to directly apply because:

- There is no differential equation whose "residual" could control the error in the sense of residual method.
- The applicability of averaging (post–processing) is based on higher regularity of exact solutions that implies the superconvergence phenomenon. Typically, solutions of variational inequalities have bounded regularity and, therefore, we cannot await such type effects.

Below we show that by the functional method it is possible to derive a posteriori estimates of the difference between the exact solution of an obstacle problem and any conforming approximation. This estimate does not require a priori knowledge on the configuration of the coincidence set.

Basic deviation estimate for variational inequalities

Let $\mathbf{a} : \mathbf{V} \times \mathbf{V} \longrightarrow \mathbb{R}$ be a bilinear $\mathbf{V}$-elliptic form and $\mathbf{j} : \mathbf{V} \longrightarrow \mathbb{R}$ be a given convex continuous functional.

Consider the following problem: find $\mathbf{u} \in \mathbf{K}$ such that the inequality

$$\mathbf{a}(\mathbf{u}, \mathbf{w} - \mathbf{u}) + \mathbf{j}(\mathbf{w}) - \mathbf{j}(\mathbf{u}) \geq \langle \mathbf{f}, \mathbf{w} - \mathbf{u} \rangle \qquad (11.3)$$

holds for any $\mathbf{w} \in \mathbf{K}$, where $\mathbf{K}$ is a convex closed subset of $\mathbf{V}$ and $\mathbf{f} \in \mathbf{V}^*$.

The solution $\mathbf{u}$ of (11.3 is a minimizer of the following variational problem $\mathcal{P}$: Find $\mathbf{u} \in \mathbf{K}$ such that

$$\mathbf{J}(\mathbf{u}) = \inf_{\mathbf{w} \in \mathbf{K}} \mathbf{J}(\mathbf{w}), \tag{11.4}$$

$$\mathbf{J}(\mathbf{w}) = \frac{1}{2}\mathbf{a}(\mathbf{w}, \mathbf{w}) + \mathbf{j}(\mathbf{w}) - \langle \mathbf{f}, \mathbf{w} \rangle .$$

Our aim is to derive a computable upper bound for the quantity $\frac{1}{2}\mathbf{a}(\mathbf{u} - \mathbf{v}, \mathbf{u} - \mathbf{v})$ where $\mathbf{v}$ is any element of the set $\mathbf{K}$.

Further analysis follows the lines of the paper

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.

First, we use (11.3) to obtain the inequality

$$\mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) = \frac{1}{2}\mathbf{a}(\mathbf{v}-\mathbf{u}, \mathbf{v}-\mathbf{u}) + \mathbf{a}(\mathbf{u}, \mathbf{v}-\mathbf{u}) - \langle \mathbf{f}, \mathbf{v}-\mathbf{u}\rangle + \\ + \mathbf{j}(\mathbf{v}) - \mathbf{j}(\mathbf{u}) \geq \frac{1}{2}\mathbf{a}(\mathbf{v}-\mathbf{u}, \mathbf{v}-\mathbf{u}),$$

which implies the **basic deviation estimate**.

$$\boxed{\frac{1}{2} \| \mathbf{v}-\mathbf{u} \|^2 \leq \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}), \quad \forall \mathbf{v} \in \mathbf{K}\,,} \tag{11.5}$$

where $\| \mathbf{w} \|^2 := \mathbf{a}(\mathbf{v}, \mathbf{v})$.

For linear problems we have derived deviation estimates by means of the inequality

$$\frac{1}{2} \| \mathbf{v} - \mathbf{u} \|^2 \leq \mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u}) = \mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{p}^*).$$

In Lectures 4 and 5 we have shown how to find a directly computable and physically meaningful upper bound of $\mathbf{J}(\mathbf{v}) - \mathbf{I}^*(\mathbf{q}^*)$.

For variational inequalities, deviation estimates are obtained in a similar way, but with some complications caused by the fact that the problem dual to $\mathcal{P}$ has a more cumbersome form.

Below, we show how we can circumvent this difficulty by using the so-called **perturbed functionals**.

Problems with two obstacles

Consider a bilinear form $\mathbf{a} : \mathbf{V_0} \times \mathbf{V_0} \longrightarrow \mathbb{R}$ defined by the relation

$$\mathbf{a}(\mathbf{v}, \mathbf{w}) := \int_\Omega \mathbf{A} \nabla \mathbf{v} \cdot \nabla \mathbf{w} \, \mathbf{dx}\,, \tag{11.6}$$

where $\mathbf{\Omega}$ is a bounded domain in $\mathbb{R}^2$ with Lipschitz continuous boundary $\partial\mathbf{\Omega}$, $\mathbf{V_0} := \overset{\circ}{\mathbf{H}}{}^1(\mathbf{\Omega})$, and $\mathbf{A} = \{\mathbf{a_{ij}}\}$ is a symmetric matrix satisfying the conditions

$$\boldsymbol{\nu}_1 \, |\xi|^2 \, \leq \, \mathbf{A}\xi \cdot \xi \, \leq \, \boldsymbol{\nu}_2 \, |\xi|^2 \,, \quad \forall \xi \in \mathbb{R}^{\mathbf{n}} \,, \quad \boldsymbol{\nu}_2 \geq \boldsymbol{\nu}_1 > \mathbf{0}.$$

Let $\mathbf{K} = \mathbf{K_{fp}} := \{\mathbf{v} \in \mathbf{V_0} \mid \varphi(\mathbf{x}) \le \mathbf{v}(\mathbf{x}) \le \psi(\mathbf{x}) \text{ a.e. in } \boldsymbol{\Omega}\}$, where $\varphi, \psi \in \mathbf{H^2}(\boldsymbol{\Omega})$ are two given functions such that

$$\varphi(\mathbf{x}) \le \psi(\mathbf{x}), \quad \forall \mathbf{x} \in \boldsymbol{\Omega},$$

Set in the general setting

$$\mathbf{j} \equiv \mathbf{0} \quad \text{and} \quad \langle \mathbf{f}, \mathbf{v} \rangle = \int_{\boldsymbol{\Omega}} \mathbf{f}\mathbf{v}\, \mathbf{dx}.$$

Then Problem $\mathcal{P}$ is the classical obstacle problem. A solution $\mathbf{u}$ minimizes the functional

$$\mathbf{J}(\mathbf{v}) = \int_{\boldsymbol{\Omega}} \mathbf{A}\nabla\mathbf{v} \cdot \nabla\mathbf{v}\mathbf{dx} - \int_{\boldsymbol{\Omega}} \mathbf{f}\mathbf{v}\mathbf{dx} \qquad \text{on } \mathbf{K_{fp}}.$$

In general, $\mathbf{\Omega}$ is devided into three sets:

$$\begin{aligned}
\mathbf{\Omega}^{\mathbf{u}}_{\oplus} &:= \quad \{\mathbf{x} \in \mathbf{\Omega} \mid \qquad \mathbf{u}(\mathbf{x}) = \psi(\mathbf{x})\} \quad (\text{upper coincidence set})\,,\\
\mathbf{\Omega}^{\mathbf{u}}_{\ominus} &:= \quad \{\mathbf{x} \in \mathbf{\Omega} \mid \qquad \mathbf{u}(\mathbf{x}) = \varphi(\mathbf{x})\} \quad (\text{lower coincidence set})\,,\\
\mathbf{\Omega}^{\mathbf{u}}_{\mathbf{0}} &:= \quad \{\mathbf{x} \in \mathbf{\Omega} \mid \varphi(\mathbf{x}) < \mathbf{u}(\mathbf{x}) < \psi(\mathbf{x})\}\ .
\end{aligned}$$

Here, $\mathbf{\Omega}^{\mathbf{u}}_{\mathbf{0}}$ is an open set, where a solution satisfies the differential equation. Thus, we see that this problem involves free boundaries, which are *unknown a priori*.
Differentiability properties of solutions to linear and quasiliniear problems with obstacles were investigated by many authors. In particular, it was proved that, under natural assumptions on external data $\mathbf{u} \in \mathbf{H}^{\mathbf{2}}(\mathbf{\Omega})$ and even for very smooth data, solutions have a limited regularity (which is $\mathbf{W}^{\mathbf{2},\infty}$). We assume that these assumptions are fulfilled and the solutions are $\mathbf{H}^{\mathbf{2}}$-regular.

Perturbed problem

To estimate the difference $\mathbf{J}(\mathbf{v}) - \mathbf{J}(\mathbf{u})$ we introduce the *perturbed functional*

$$\mathbf{J}_{\boldsymbol{\lambda}}(\mathbf{v}) := \mathbf{J}(\mathbf{v}) - \int_{\boldsymbol{\Omega}} \boldsymbol{\lambda} \cdot (\upsilon - \boldsymbol{\Phi})\, \mathbf{dx},$$

where $\boldsymbol{\Phi} = (\varphi, -\psi)$ and $\upsilon = (\,\mathbf{v},\, -\mathbf{v})$,

$$\boldsymbol{\lambda} \in \aleph_{\oplus} := \left\{ (\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \mid \boldsymbol{\lambda}_\mathbf{i} \in \mathbf{L}^2(\boldsymbol{\Omega}),\ \boldsymbol{\lambda}_\mathbf{i}(\mathbf{x}) \geq \mathbf{0}\, \text{a.e. in } \boldsymbol{\Omega},\ \mathbf{i} = \mathbf{1}, \mathbf{2} \right\}.$$

It is easy to see that

$$\begin{aligned} \sup_{\boldsymbol{\lambda} \in \aleph_{\oplus}} \mathbf{J}_{\boldsymbol{\lambda}}(\mathbf{v}) &= \mathbf{J}(\mathbf{v}) - \inf_{\boldsymbol{\lambda} \in \aleph_{\oplus}} \int_{\boldsymbol{\Omega}} \boldsymbol{\lambda} \cdot (\upsilon - \boldsymbol{\Phi})\, \mathbf{dx} \\ &= \begin{cases} \mathbf{J}(\mathbf{v}) & \text{if } \mathbf{v} \in \mathbf{K}_{\mathbf{fp}}, \\ +\infty & \text{if } \mathbf{v} \notin \mathbf{K}_{\mathbf{fp}}. \end{cases} \end{aligned} \tag{11.7}$$

With $\mathbf{J}_{\boldsymbol{\lambda}}$ we associate a *perturbed variational problem*.
Problem $\mathcal{P}_{\boldsymbol{\lambda}}$. Find $\mathbf{u}_{\boldsymbol{\lambda}} \in \mathbf{V_0}$ such that

$$\mathbf{J}_{\boldsymbol{\lambda}}(\mathbf{u}_{\boldsymbol{\lambda}}) = \inf_{\mathbf{v} \in \mathbf{V_0}} \mathbf{J}_{\boldsymbol{\lambda}}(\mathbf{v}) := \inf \mathcal{P}_{\boldsymbol{\lambda}}. \tag{11.8}$$

Since

$$\inf_{\mathbf{v} \in \mathbf{V_0}} \mathbf{J}_{\boldsymbol{\lambda}}(\mathbf{v}) \leq \inf_{\mathbf{v} \in \mathbf{K_{fp}}} \mathbf{J}_{\boldsymbol{\lambda}}(\mathbf{v}) = \inf_{\mathbf{v} \in \mathbf{K_{fp}}} \mathbf{J}(\mathbf{v}) = \inf \mathcal{P},$$

we see that

$$\inf \mathcal{P}_{\boldsymbol{\lambda}} \leq \inf \mathcal{P}, \quad \forall \boldsymbol{\lambda} \in \aleph_{\oplus}. \tag{11.9}$$

It follows from (11.5) and (11.9) that

$$\frac{1}{2} \,|\!|\!| \mathbf{v} - \mathbf{u} |\!|\!|^2 \leq \mathbf{J}(\mathbf{v}) - \inf \mathcal{P}_{\boldsymbol{\lambda}}, \qquad \boldsymbol{\lambda} \in \aleph_{\oplus}. \quad (11.10)$$

To estimate the right-hand side of (11.9), we introduce a dual counterpart of Problem $\mathcal{P}_{\boldsymbol{\lambda}}$.

Dual perturbed problem

By the Lagrangian

$$\mathbf{L}(\mathbf{v},\boldsymbol{\tau},\boldsymbol{\lambda}) := \int_\Omega \left(\boldsymbol{\tau}\cdot\nabla\mathbf{v} - \frac{1}{2}\mathbf{A}^{-1}\boldsymbol{\tau}\cdot\boldsymbol{\tau} - \mathbf{f}\mathbf{v} - \boldsymbol{\lambda}\cdot(v-\boldsymbol{\Phi}) \right) \mathbf{dx},$$

we define the perturbed functional as follows:

$$\mathbf{J}_{\boldsymbol{\lambda}}(\mathbf{v}) = \sup_{\boldsymbol{\tau}\in\mathbf{Y}^*} \mathbf{L}(\mathbf{v},\boldsymbol{\tau},\boldsymbol{\lambda}), \qquad \mathbf{Y}^* := \mathbf{L}^2(\boldsymbol{\Omega},\mathbb{R}^2).$$

Problem $\mathcal{P}^*_{\boldsymbol{\lambda}}$. Find $\boldsymbol{\tau}_{\boldsymbol{\lambda}}$ such that

$$\mathbf{J}^*_{\boldsymbol{\lambda}}(\boldsymbol{\tau}_{\boldsymbol{\lambda}}) = \sup_{\mathbf{q}\in\mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}}} \mathbf{J}^*_{\boldsymbol{\lambda}}(\mathbf{q}),$$

where $\mathbf{J}^*_{\boldsymbol{\lambda}}(\mathbf{q}) = \int_\Omega \left(-\frac{1}{2}\mathbf{A}^{-1}\mathbf{q}\cdot\mathbf{q} + \boldsymbol{\lambda}\cdot\boldsymbol{\Phi}\right) \mathbf{dx}$ and

$$\mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}} := \left\{ \mathbf{q}\in\mathbf{Y}^* \middle| \int_\Omega \mathbf{q}\cdot\nabla\mathbf{v}\,\mathbf{dx} = \int_\Omega (\mathbf{f}\mathbf{v} + \boldsymbol{\lambda}\cdot v)\,\mathbf{dx}\ \forall\mathbf{v}\in\mathbf{V_0} \right\}.$$

$\mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}}$ is a closed affine manifold in $\mathbf{Y}^*$ and the functional $-\mathbf{J}^*_{\boldsymbol{\lambda}}$ is convex and continuous on $\mathbf{Y}^*$.
Therefore, Problem $\mathcal{P}^*_{\boldsymbol{\lambda}}$ has a solution and

$$\inf \mathcal{P}_{\boldsymbol{\lambda}} = \sup \mathcal{P}^*_{\boldsymbol{\lambda}}. \tag{11.11}$$

Estimates of the deviation

By means of (11.5) and (11.11) we obtain

$$\frac{1}{2} \| \mathbf{v} - \mathbf{u} \|^2 \leq \mathbf{J}(\mathbf{v}) - \sup \mathcal{P}^*_{\boldsymbol{\lambda}} \leq \mathbf{J}(\mathbf{v}) - \mathbf{J}^*_{\boldsymbol{\lambda}}(\mathbf{q}) \tag{11.12}$$

Here

$$\mathbf{v} \in \mathbf{K_{fp}}, \quad \mathbf{q} \in \mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}}, \quad \boldsymbol{\lambda} \in \aleph_{\oplus}.$$

Rewrite $\mathbf{J}(\mathbf{v}) - \mathbf{J}^*_{\boldsymbol{\lambda}}(\mathbf{q})$ in a more transparent form. We have

$$\frac{1}{2} \| \mathbf{v} - \mathbf{u} \|^2 \leq \int_\Omega \left(\frac{1}{2} \mathbf{A} \nabla \mathbf{v} \cdot \nabla \mathbf{v} - \mathbf{f}\mathbf{v} \right) \mathbf{dx} + \int_\Omega \frac{1}{2} \mathbf{A}^{-1} \boldsymbol{\tau} \cdot \boldsymbol{\tau} \, \mathbf{dx}$$
$$+ \frac{1}{2} \int_\Omega \left(\mathbf{A}^{-1} \mathbf{q} \cdot \mathbf{q} - \mathbf{A}^{-1} \boldsymbol{\tau} \cdot \boldsymbol{\tau} \right) \mathbf{dx} - \int_\Omega \boldsymbol{\lambda} \cdot \boldsymbol{\Phi} \, \mathbf{dx}.$$

In view of the relation

$$\mathbf{A}^{-1}\mathbf{a}\cdot\mathbf{a}-\mathbf{A}^{-1}\mathbf{b}\cdot\mathbf{b}=\mathbf{A}^{-1}(\mathbf{a}-\mathbf{b})\cdot(\mathbf{a}-\mathbf{b})+2\mathbf{A}^{-1}\mathbf{b}\cdot(\mathbf{a}-\mathbf{b})$$

and the integral identity

$$\int_\Omega \mathbf{f}\mathbf{v}\,\mathbf{dx}=\int_\Omega \mathbf{q}\cdot\nabla\mathbf{v}\,\mathbf{dx}-\int_\Omega \boldsymbol{\lambda}\cdot\upsilon\,\mathbf{dx},\quad \forall\mathbf{q}\in\mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}},$$

we obtain

$$\begin{aligned}\frac{1}{2}\|\,\mathbf{v}-\mathbf{u}\,\|^2\leq \frac{1}{2}\int_\Omega(\mathbf{A}\nabla\mathbf{v}-\boldsymbol{\tau})\cdot(\nabla\mathbf{v}-\mathbf{A}^{-1}\boldsymbol{\tau})\,\mathbf{dx}\\ +\int_\Omega\boldsymbol{\lambda}\cdot(\upsilon-\boldsymbol{\Phi})\mathbf{dx}+\frac{1}{2}\int_\Omega\mathbf{A}^{-1}(\mathbf{q}-\boldsymbol{\tau})\cdot(\mathbf{q}-\boldsymbol{\tau})\\ +\int_\Omega(\nabla\mathbf{v}-\mathbf{A}^{-1}\boldsymbol{\tau})\cdot(\boldsymbol{\tau}-\mathbf{q})\,\mathbf{dx}.\end{aligned}$$

The last integral can be estimated as follows:

$$\int_{\Omega} (\nabla \mathbf{v} - \mathbf{A}^{-1}\boldsymbol{\tau}) \cdot (\boldsymbol{\tau} - \mathbf{q}) \mathbf{dx}$$

$$\leq \frac{\beta}{2} \int_{\Omega} \mathbf{A}(\nabla \mathbf{v} - \mathbf{A}^{-1}\boldsymbol{\tau}) \cdot (\nabla \mathbf{v} - \mathbf{A}^{-1}\boldsymbol{\tau})\, \mathbf{dx}$$

$$+ \frac{1}{2\beta} \int_{\Omega} \mathbf{A}^{-1}(\mathbf{q} - \boldsymbol{\tau}) \cdot (\mathbf{q} - \boldsymbol{\tau})\, \mathbf{dx},$$

where β is any positive number.
Introduce the quantity

$$\mathbf{d}^2(\boldsymbol{\tau}, \mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}}) := \inf_{\mathbf{q} \in \mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}}} \int_{\Omega} \mathbf{A}^{-1}(\mathbf{q} - \boldsymbol{\tau}) \cdot (\mathbf{q} - \boldsymbol{\tau})\, \mathbf{dx}\,,$$

which is the distance between $\boldsymbol{\tau}$ and the set $\mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}}$.

Now, we rewrite the estimate as follows:

$$\| \mathbf{v} - \mathbf{u} \|^2 \leq$$

$$\leq (1+\beta)\mathbf{d}^2(\boldsymbol{\tau}, \mathbf{Q}^*_{\mathbf{f}\boldsymbol{\lambda}}) + \left(1 + \frac{1}{\beta}\right) \| \nabla \mathbf{v} - \mathbf{A}^{-1}\boldsymbol{\tau} \|^2 +$$

$$+ 2\int_{\Omega} \boldsymbol{\lambda} \cdot (v - \boldsymbol{\Phi})\, \mathbf{dx}. \quad (11.13)$$

Recall some relations that has been established in Lecture 5. We have proved that

$$\mathbf{d}(\mathbf{y}, \mathbf{Q}^*_\ell) = [\![\, \ell + \mathbf{\Lambda}^* \mathbf{y} \,]\!] := \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\langle \ell + \mathbf{\Lambda}^* \mathbf{y}, \mathbf{w} \rangle}{\|\, \mathbf{\Lambda w} \,\|},$$

where $\mathbf{Q}^*_\ell := \{\mathbf{y} \in \mathbf{Y}^* \mid (\mathbf{y}, \mathbf{\Lambda w}) + \langle \ell, \mathbf{w} \rangle = \mathbf{0}, \quad \forall \mathbf{w} \in \mathbf{V_0}\}$.
In our case, $\mathbf{y} = \boldsymbol{\tau}$, $\mathbf{\Lambda} = \nabla$, $\mathbf{\Lambda}^* = -\mathbf{div}$, and $Q^*_\ell = Q^*_{f\boldsymbol{\lambda}}$ if set

$$\langle \ell, \mathbf{w} \rangle = -\int_{\Omega} (\mathbf{fw} + \boldsymbol{\lambda} \cdot v) \mathbf{dx}.$$

Therefore,

$$\mathbf{d}(\mathbf{y}, \mathbf{Q}^*_\ell) = \sup_{\mathbf{w} \in \mathbf{V_0}} \frac{\int_{\Omega} (-\mathbf{f} - \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 - \mathbf{divy}) \mathbf{wdx}}{\|\, \mathbf{w} \,\|}.$$

Assume that $\boldsymbol{\tau} \in \mathbf{Q}^* := \mathbf{H}(\boldsymbol{\Omega}, \mathrm{div})$. Then,

$$\int_{\boldsymbol{\Omega}} (-\mathbf{f} - \boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_2 - \mathbf{div y})\mathbf{w}\mathbf{dx} \leq \mathbf{C}_{\boldsymbol{\Omega},\mathbf{A}} \|\mathbf{f} + \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2 + \mathbf{div y}\| \; |\!|\!| \; \mathbf{w} \; |\!|\!|$$

and we obtain

$$\mathbf{d}(\boldsymbol{\tau}, \mathbf{Q}^*_{\boldsymbol{\lambda}}) \leq \mathbf{C}_{\boldsymbol{\Omega},\mathbf{A}} \left\| \mathbf{div}\,\boldsymbol{\tau} + \mathbf{f} + \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2 \right\| , \qquad (11.14)$$

where $\mathbf{C}_{\boldsymbol{\Omega},\mathbf{A}}$ is a constant in the inequality

$$\|\mathbf{w}\|_{\mathbf{2},\boldsymbol{\Omega}} \leq \mathbf{C}_{\boldsymbol{\Omega},\mathbf{A}} \; |\!|\!| \; \mathbf{w} \; |\!|\!|, \quad \forall \mathbf{w} \in \mathbf{V_0}.$$

Thus, for $\mathbf{y}^* \in \mathbf{Q}^*$ we obtain the estimate

$$\|\!| \mathbf{v} - \mathbf{u} \|\!|^2 \leq \mathbf{C}^2_{\boldsymbol{\Omega},\mathbf{A}}(\mathbf{1}+\beta)\,\|\mathbf{div}\,\boldsymbol{\tau}+\mathbf{f}+\boldsymbol{\lambda}_1-\boldsymbol{\lambda}_2\|^2$$

$$+\left(\mathbf{1}+\frac{\mathbf{1}}{\beta}\right)\|\!| \nabla\mathbf{v}-\mathbf{A}^{-1}\boldsymbol{\tau}\|\!|^2 + 2\int_{\boldsymbol{\Omega}}\boldsymbol{\lambda}\cdot(v-\boldsymbol{\Phi})\,\mathbf{dx}. \quad (11.15)$$

In this estimate, $\boldsymbol{\lambda}$ is a "free" vector–valued function. We use this freedom to obtain the most accurate upper bound for the deviation

$$\|\!| \mathbf{v} - \mathbf{u} \|\!|.$$

Below we consider two options that lead to two different a posteriori error estimates for the obstacle problem.

The first option is as follows. Let $\mathbf{v} \in \mathbf{V}$ be an approximate solution. For almost all points of $\mathbf{\Omega}$ the function $\mathbf{v}(\mathbf{x})$ is either equal to φ, or ψ or lies between these two values. Thus, almost all points of $\mathbf{\Omega}$ can be referred to one of the three sets:

$$
\begin{aligned}
\mathbf{\Omega}_0^{\mathbf{v}} &:= \ \{\mathbf{x} \in \mathbf{\Omega} \mid \varphi(\mathbf{x}) < \mathbf{v}(\mathbf{x}) < \psi(\mathbf{x})\}\,, \\
\mathbf{\Omega}_\ominus^{\mathbf{v}} &:= \ \{\mathbf{x} \in \mathbf{\Omega} \mid \quad \mathbf{v}(\mathbf{x}) = \varphi(\mathbf{x})\}\,, \\
\mathbf{\Omega}_\oplus^{\mathbf{v}} &:= \ \{\mathbf{x} \in \mathbf{\Omega} \mid \quad \mathbf{v}(\mathbf{x}) = \psi(\mathbf{x})\}\,.
\end{aligned}
$$

Now, we can choose $\boldsymbol{\lambda}$ as follows:

$$
\begin{aligned}
\boldsymbol{\lambda}_1 = \boldsymbol{\lambda}_2 &= \mathbf{0} && \text{a.e. in } \mathbf{\Omega}_0^{\mathbf{v}}, \\
\boldsymbol{\lambda}_1 = -\langle\, \mathbf{div}\, \boldsymbol{\tau} + \mathbf{f}\,\rangle_\ominus,\ \boldsymbol{\lambda}_2 &= \mathbf{0} && \text{a.e. in } \mathbf{\Omega}_\ominus^{\mathbf{v}}, \\
\boldsymbol{\lambda}_1 = \mathbf{0},\ \boldsymbol{\lambda}_2 &= \langle\, \mathbf{div}\, \boldsymbol{\tau} + \mathbf{f}\,\rangle_\oplus, && \text{a.e. in } \mathbf{\Omega}_\oplus^{\mathbf{v}}.
\end{aligned}
$$

Here $\langle\, \mathbf{z}\,\rangle_\oplus$ is zero if $\mathbf{z} \leq \mathbf{0}$ and $\mathbf{z}$ if $\mathbf{z} > \mathbf{0}$.

As a result of such a choice of $\boldsymbol{\lambda}$, we obtain the estimate

$$\| \mathbf{v} - \mathbf{u} \|^2 \leq \mathbf{M_1}(\mathbf{v}, \boldsymbol{\tau}, \beta) := \left(1 + \frac{1}{\beta}\right) \| \nabla \mathbf{v} - \mathbf{A}^{-1}\boldsymbol{\tau} \|^2$$

$$+ \mathbf{C}^2_{\Omega,\mathbf{A}}(1+\beta)\left[\int\limits_{\Omega_0^{\mathbf{v}}} |\mathbf{r}(\boldsymbol{\tau})|^2 \, \mathbf{dx} + \int\limits_{\Omega_\oplus^{\mathbf{v}}} \langle \mathbf{r}(\boldsymbol{\tau}) \rangle^2_\ominus \, \mathbf{dx} + \int\limits_{\Omega_\ominus^{\mathbf{v}}} \langle \mathbf{r}(\boldsymbol{\tau}) \rangle^2_\oplus \, \mathbf{dx} \right], \tag{11.16}$$

where

$$\mathbf{r}(\boldsymbol{\tau}) = \mathbf{div}\, \boldsymbol{\tau} + \mathbf{f}$$

and $\langle \ \rangle_\ominus$ and $\langle \ \rangle_\oplus$ denote the negative and positive parts of a quantity, respectively.

What is the meaning of the four terms of the Majorant?

The first term $\| \nabla \mathbf{v} - \mathbf{A}^{-1} \boldsymbol{\tau} \|^2$ **penalizes the error in the duality relation**

$$\nabla \mathbf{v} = \mathbf{A}^{-1} \boldsymbol{\tau}.$$

Other terms **penalize "improper" behavior** of $\mathbf{r}(\boldsymbol{\tau})$ on the sets $\boldsymbol{\Omega}_{\mathbf{0}}^{\mathbf{v}}$, $\boldsymbol{\Omega}_{\oplus}^{\mathbf{v}}$, and $\boldsymbol{\Omega}_{\ominus}^{\mathbf{v}}$, respectively. Indeed, on $\boldsymbol{\Omega}_{\mathbf{0}}^{\mathbf{v}}$ the differential equation must be satisfied. Therefore, the term

$$\int_{\boldsymbol{\Omega}_{\mathbf{0}}^{\mathbf{v}}} |\mathbf{r}(\boldsymbol{\tau})|^2 \, \mathbf{dx}$$

can be viewed as a penalty, which is nonzero if the variable $\boldsymbol{\tau}$ (flux image) does not satisfy the differential equation.

By the necessary conditions for the obstacle problem, we find that

$$\begin{aligned} &\mathbf{div}\, A\nabla \mathbf{u}(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \leq \mathbf{0}, && \text{for a. e. } \mathbf{x} \in \mathbf{\Omega}^{\mathbf{v}}_{\ominus}, \\ &\mathbf{div}\, A\nabla \mathbf{u} + \mathbf{f}(\mathbf{x}) \geq \mathbf{0} && \text{for a. e. } \mathbf{x} \in \mathbf{\Omega}^{\mathbf{v}}_{\oplus}. \end{aligned}$$

Thus, the terms

$$\int\limits_{\mathbf{\Omega}^{\mathbf{v}}_{\oplus}} \langle\, \mathbf{r}(\tau)\,\rangle^{2}_{\ominus} \mathbf{dx} \quad \text{and} \quad \int\limits_{\mathbf{\Omega}^{\mathbf{v}}_{\ominus}} \langle\, \mathbf{r}(\tau)\,\rangle^{2}_{\oplus} \mathbf{dx}$$

are certain penalties for the violation of above conditions.

We see that the majorant $\mathbf{M_1}$ is a nonnegative functional, which vanishes if and only if

$$\nabla \mathbf{v}(\mathbf{x}) = \mathbf{A}^{-1}\boldsymbol{\tau}(\mathbf{x}) \qquad \text{for a. e. } \mathbf{x} \in \boldsymbol{\Omega}, \tag{11.17}$$

$$\mathbf{div}\,\boldsymbol{\tau}(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \leq \mathbf{0}, \qquad \text{for a. e. } \mathbf{x} \in \boldsymbol{\Omega}^{\mathbf{v}}_{\ominus}, \tag{11.18}$$

$$\mathbf{div}\,\boldsymbol{\tau}(\mathbf{x}) + \mathbf{f}(\mathbf{x}) = \mathbf{0}, \qquad \text{for a. e. } \mathbf{x} \in \boldsymbol{\Omega}^{\mathbf{v}}_{\mathbf{0}}, \tag{11.19}$$

$$\mathbf{div}\,\boldsymbol{\tau}(\mathbf{x}) + \mathbf{f}(\mathbf{x}) \geq \mathbf{0} \qquad \text{for a. e. } \mathbf{x} \in \boldsymbol{\Omega}^{\mathbf{v}}_{\oplus}. \tag{11.20}$$

Let us show that in this case $\mathbf{v} = \mathbf{u}$ and $\boldsymbol{\tau} = \mathbf{A}\nabla\mathbf{u}$.

Assume that (11.17)–(11.20) hold. Then for any $\mathbf{w} \in \mathbf{K_{fp}}$, we have

$$\int_{\Omega} \mathbf{A}\nabla \mathbf{v} \cdot \nabla(\mathbf{w}-\mathbf{v})\,\mathbf{dx} - \int_{\Omega} \mathbf{f}(\mathbf{w}-\mathbf{v})\,\mathbf{dx}$$
$$= \int_{\Omega} (\mathbf{div}\,\boldsymbol{\tau} + \mathbf{f})(\mathbf{v}-\mathbf{w})\,\mathbf{dx} = \int_{\Omega^{\mathbf{v}}_{\ominus}} (\mathbf{div}\,\boldsymbol{\tau} + \mathbf{f})(\varphi - \mathbf{w})\,\mathbf{dx}$$
$$+ \int_{\Omega^{\mathbf{v}}_{0}} (\mathbf{div}\,\boldsymbol{\tau} + \mathbf{f})(\mathbf{v}-\mathbf{w})\,\mathbf{dx} + \int_{\Omega^{\mathbf{v}}_{\oplus}} (\mathbf{div}\,\boldsymbol{\tau} + \mathbf{f})(\psi - \mathbf{w})\,\mathbf{dx} \geq \mathbf{0}.$$

This inequality means that

$$\mathbf{a}(\mathbf{v}, \mathbf{w}-\mathbf{v}) \geq \int_{\Omega} \mathbf{f}(\mathbf{w}-\mathbf{v})\,\mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{K_{fp}}\,,$$

so that $\mathbf{v}$ coincides with the exact solution $\mathbf{u}$ (which is unique!).

All said above can be summarized as follows:

Theorem

For any $\beta > \mathbf{0}$, $\mathbf{M_1}(\mathbf{v}, \boldsymbol{\tau}, \beta)$ *is a nonnegative functional that majorizes* $\| \mathbf{v} - \mathbf{u} \|^2$ *and vanishes if and only if*

$$\mathbf{v} = \mathbf{u} \quad \text{and} \quad \boldsymbol{\tau} = \mathbf{A}\nabla\mathbf{u},$$

where $\mathbf{u}$ *is a solution of the variational inequality*

$$\mathbf{a}(\mathbf{u}, \mathbf{w} - \mathbf{u}) \geq \int_{\Omega} \mathbf{f}(\mathbf{w} - \mathbf{u})\,\mathbf{dx}, \quad \forall \mathbf{w} \in \mathbf{K_{fp}}\,,$$

To obtain a more rigorous upper bound of $\| \mathbf{v} - \mathbf{u} \|$, we should find $\boldsymbol{\lambda}$ by minimizing the right-hand side of the estimate

$$\| \mathbf{v} - \mathbf{u} \|^2 \leq \mathbf{C}^2_{\boldsymbol{\Omega},\mathbf{A}}(\mathbf{1} + \beta) \, \|\mathbf{div}\, \boldsymbol{\tau} + \mathbf{f} + \boldsymbol{\lambda}_1 - \boldsymbol{\lambda}_2\|^2 + \left(\mathbf{1} + \frac{\mathbf{1}}{\beta}\right) \| \nabla \mathbf{v} - \mathbf{A}^{-1}\boldsymbol{\tau} \|^2 + \mathbf{2}\int_{\boldsymbol{\Omega}} \boldsymbol{\lambda}\cdot(v - \boldsymbol{\Phi})\, \mathbf{dx}.$$

Note that it leads to a quadratic type minimization problem in $\mathbf{L}^2$ that can be solved analytically. On this way, we arrive at the estimate

$$\| \mathbf{v} - \mathbf{u} \|^2 \leq \mathbf{M}_2(\mathbf{v}, \boldsymbol{\tau}, \beta) := (\mathbf{1} + \frac{\mathbf{1}}{\beta}) \| \nabla \mathbf{v} - \mathbf{A}^{-1}\boldsymbol{\tau} \|^2 + \int_{\boldsymbol{\Omega}} \mathbf{R}(\mathbf{v}, \mathbf{div}\, \boldsymbol{\tau} + \mathbf{f}, \beta)\, \mathbf{dx}\,, \quad (11.21)$$

where

$$\mathbf{R}(\mathbf{v},\mathbf{r},\beta)=\begin{cases} -\frac{(\varphi-\mathbf{v})^2}{\mathbf{c}_\beta}+2\mathbf{r}(\varphi-\mathbf{v}) & \text{if} \quad \mathbf{c}_\beta\mathbf{r}+\mathbf{v}\leq\ \varphi, \\ \mathbf{c}_\beta\mathbf{r}^2 & \text{if} \quad \varphi<\ \mathbf{c}_\beta\mathbf{r}+\mathbf{v}<\psi, \\ -\frac{(\psi-\mathbf{v})^2}{\mathbf{c}_\beta}+2\mathbf{r}(\psi-\mathbf{v}) & \text{if} \quad \mathbf{c}_\beta\mathbf{r}+\mathbf{v}\geq\psi \end{cases}$$

and $\mathbf{c}_\beta=\mathbf{C}^2_{\Omega,\mathbf{A}}(1+\beta)$.

Let us show that the term $\mathbf{R}(\mathbf{v},\mathbf{r},\beta)$ is equal to zero in the following three cases:

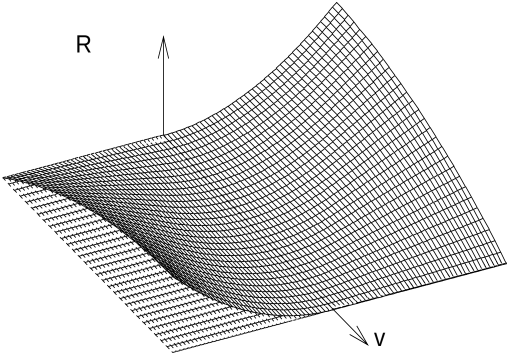
$$(\mathbf{I})\,\mathbf{r}=\mathbf{0},\qquad (\mathbf{II})\,\mathbf{v}=\varphi \text{ and } \mathbf{r}<\mathbf{0},\qquad (\mathbf{III})\,\mathbf{v}=\psi \text{ and } \mathbf{r}>\mathbf{0}.$$

Assume that $\mathbf{r}=\mathbf{0}$. If $\varphi<\mathbf{v}<\psi$, then the second branch is realized and we see that $\mathbf{R}=\mathbf{0}$. If $\varphi=\mathbf{v}$ (or $\psi=\mathbf{v}$), then the first (third) branch is realized and also $\mathbf{R}=\mathbf{0}$.
Let $\mathbf{r}>\mathbf{0}$ (the case $\mathbf{r}<\mathbf{0}$ is considered quite similar). Then the first branch is impossible. On the second one we have only positive values. For the third branch we have

$$\mathbf{r}\geq\frac{\psi-\mathbf{v}}{\mathbf{c}_\beta}\quad\text{and, therefore,}\quad -\frac{(\psi-\mathbf{v})^2}{\mathbf{c}_\beta}+\mathbf{2r}(\psi-\mathbf{v})\geq\frac{(\psi-\mathbf{v})^2}{\mathbf{c}_\beta}.$$

We see that this quantity can be zero if and only if $\mathbf{v}=\psi$.

This behavior of $\mathbf{R}(\mathbf{v}, \mathbf{r}, \beta)$ is clearly observed on the figure below, where $\varphi = \mathbf{0}$ and $\psi = \mathbf{1}$.



The functional $\mathbf{M_2}$ is defined for any $\mathbf{v} \in \mathbf{K_{fp}}$, $\boldsymbol{\tau} \in \mathbf{Q}^*$, and $\beta > \mathbf{0}$. It is clear that

$$\mathbf{M_2}(\mathbf{v}, \boldsymbol{\tau}, \beta) \,\leq\, \mathbf{M_1}(\mathbf{v}, \boldsymbol{\tau}, \beta).$$

This fact immediately implies the following assertion.

Theorem

For any $\beta > \mathbf{0}$, $\mathbf{M_2}(\mathbf{v}, \boldsymbol{\tau}, \beta)$ is a nonnegative functional that majorizes $\| \mathbf{v} - \mathbf{u} \|^{\mathbf{2}}$ and vanishes if and only if

$$\mathbf{v} = \mathbf{u} \quad \text{and} \quad \boldsymbol{\tau} = \mathbf{A}\nabla\mathbf{u},$$

where $\mathbf{u}$ is a solution of the obstacle problem.

Approximative properties

It is not difficult to prove that for any $\beta > \mathbf{0}$ the functional $\mathbf{M_2}(\mathbf{v}, \boldsymbol{\tau}, \beta)$ possesses necessary continuity properties with respect to the first and second arguments. Namely,

$$\mathbf{M_2}(\mathbf{v_k}, \boldsymbol{\tau}_\mathbf{k}, \beta) \;\longrightarrow\; \mathbf{0}$$

if

$$\mathbf{v_k} \longrightarrow \mathbf{u} \quad \text{in} \quad \mathbf{V_0}$$

and

$$\boldsymbol{\tau}_\mathbf{k} \longrightarrow \mathbf{A}\nabla\mathbf{u} \quad \text{in} \quad \mathbf{Q}^*.$$

If the problem contains only one obstacle (e.g., if $\psi = +\infty$), then the function $\mathbf{R}$ has a more compact form:

$$\mathbf{R}(\mathbf{v}, \boldsymbol{\tau}, \beta) = \mathbf{c}_{\beta}\left[|\mathbf{r}|^2 - \langle\, \frac{\mathbf{v} - \varphi}{\mathbf{c}_{\beta}} + \mathbf{r}\,\rangle_{\ominus}^{2}\right].$$

For a membrane problem, this case was analyzed in

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.

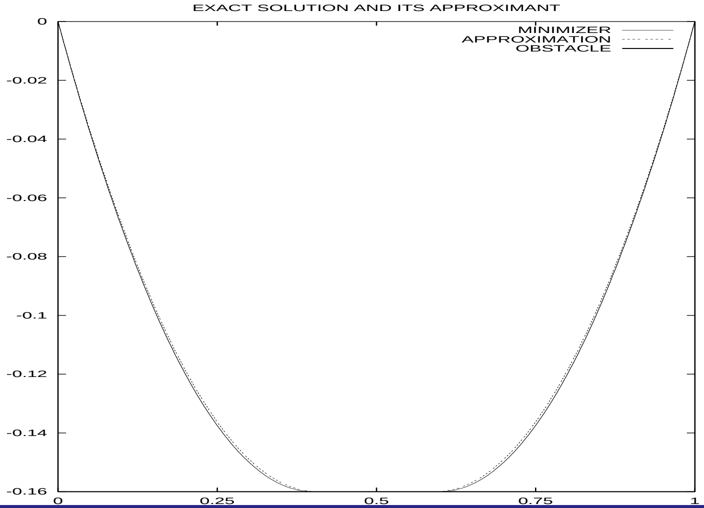
Numerical tests. Example 1

We start with simple 1D tests, where the equation is $\mathbf{u}'' = \mathbf{f}$ on $(\mathbf{0}, \mathbf{1})$ and the boundary conditions are homogeneous.
An approximate solution was computed for a uniform mesh with 60 intervals. In this example,

$$\mathbf{f} = -\mathbf{2.0}, \qquad \varphi(\mathbf{x}) = -\mathbf{0.16},$$

and the coincidence set is $[.\mathbf{400}, .\mathbf{600}]$. The minimal value of the functional is $-.\mathbf{149}$.

Exact solution, obstacle and approximate solution



In this case, the error is **0.000118**.
$\mathbf{M}_{\oplus}$ computed for $\mathbf{y}^* = \mathbf{G_h}(\nabla \mathbf{u_h})$ (when the dual variable $\mathbf{y}^*$ is computed by a simple gradient averaging procedure)
gives the first upper bound **0.000647**. Thus, without noticeable additional expenditures, we obtain an estimate with

$$\mathbf{I_{eff}} = \mathbf{5.473}.$$

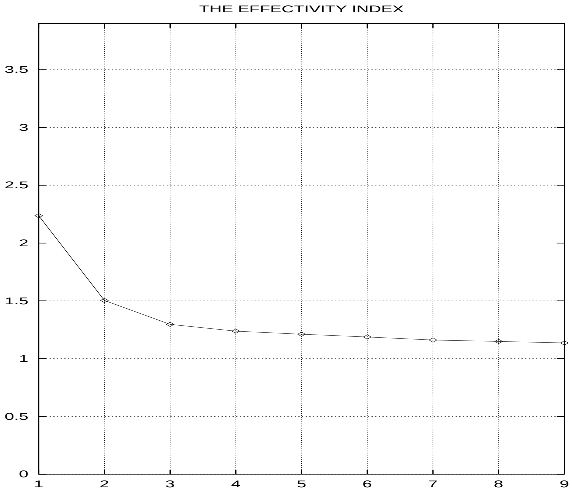
In this case, two parts of the Majorant have the following values:
0.000164 (duality term) and
0.000483 (generalized residual term).

Then, $\mathbf{M}_{\oplus}$ was minimized with respect to the dual variable. In the table, we present values of the Majorant obtained in this process. Computational expenditures are measured by the "time unit", which is the time required for computing the approximate solution.

Table:

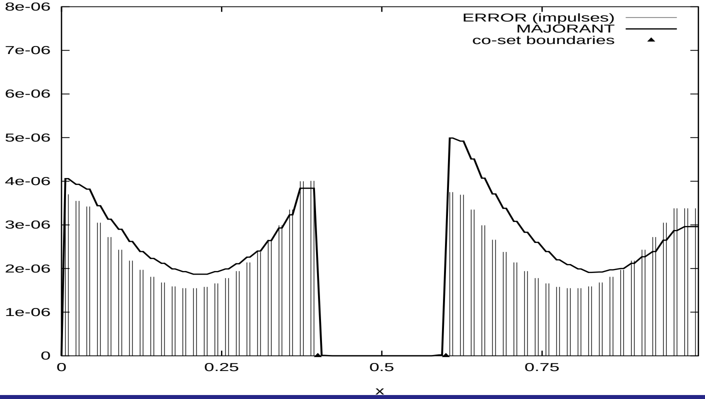
Iteration	The majorant	$\mathbf{I}_{\mathbf{eff}}$	Expenditures
1	.000214	1.804802	.448
2	.000163	1.379279	.660
3	.000152	1.281185	.787
4	.000146	1.232812	.881
5	.000143	1.209496	.977
6	.000140	1.185637	1.073
7	.000137	1.158350	1.169
8	.000136	1.148871	1.243
9	.000134	1.134931	1.336

This process is depicted below



Error indication

Distributions of subinterval errors and errors computed by the Majorant are depicted on the next picture. We see that $\mathbf{M}_\oplus$ provides a good representation of the error distribution.

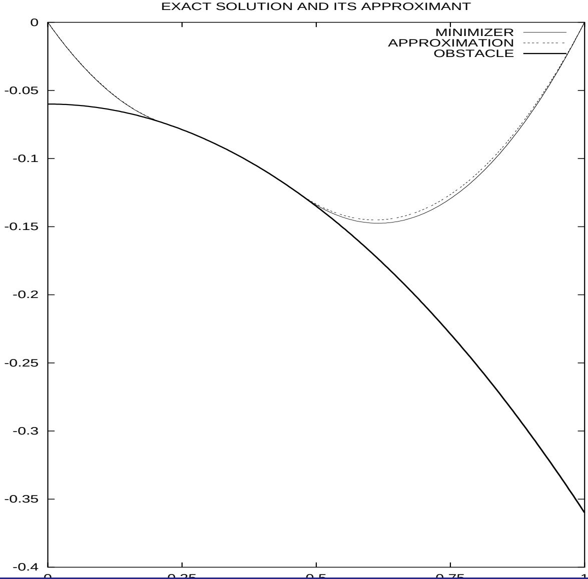


Numerical tests. Example 2

Take the same problem with $\mathbf{f} = -\mathbf{2.0}$ and

$$\varphi(\mathbf{x}) = -\mathbf{0.3x^2} - \mathbf{0.06}.$$

In this case $\mathbf{\Omega}_\varphi = [.\mathbf{215}, .\mathbf{474}]$ and the lower bound of the primal variational problem is equal to $-.\mathbf{125}$. An approximate solution was computed for the uniform mesh with 60 subintervals.


EXACT SOLUTION AND ITS APPROXIMANT
MINIMIZER
APPROXIMATION
OBSTACLE
0
-0.05
-0.1
-0.15
-0.2
-0.25
-0.3
-0.35
-0.4

In this example, the error is $\mathbf{0.000158}$. The value of $\mathbf{M}_\oplus$ computed for $\mathbf{y}^* = \mathbf{G_h}(\nabla \mathbf{u_h})$ gives the first (rough) upper bound of the error 0.000861. Thus, without serious additional expenditures, we obtain an estimate with

$$\mathbf{I_{eff}} = \mathbf{5.457}.$$

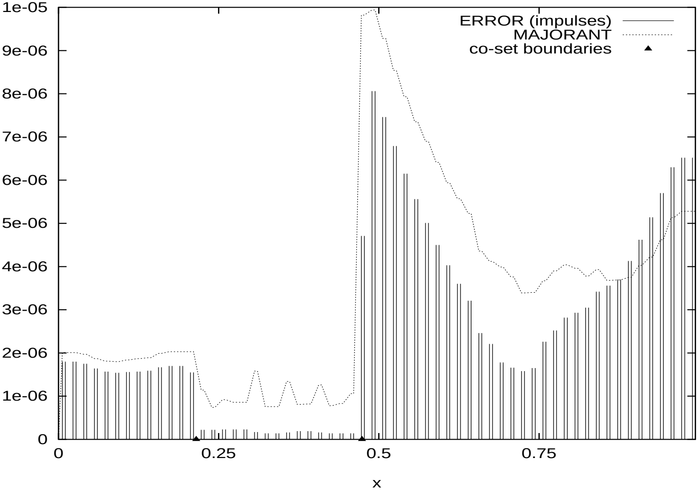
Two parts of the Majorant are as follows:
$\mathbf{0.000156}$ (duality term) and $\mathbf{0.000704}$ (generalized residual term).

Then, the Majorant was minimized with respect to $\mathbf{y}$. The respective results are presented below

Table:

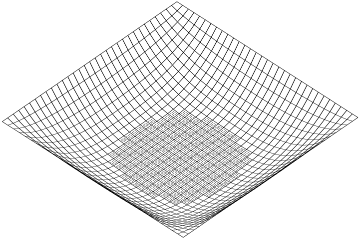
Iteration	The majorant	$\mathbf{I}_{\mathbf{eff}}$	Expenditures
1	.000296	1.877766	.625
2	.000245	1.551655	.842
3	.000240	1.521251	.938
4	.000235	1.492785	1.033
5	.000227	1.439252	1.165
6	.000218	1.379380	1.287
7	.000212	1.343913	1.383
8	.000210	1.334278	1.457
9	.000209	1.322333	1.542

In the next figure, we show the distribution of actual errors on the intervals and those computed by the Majorant.

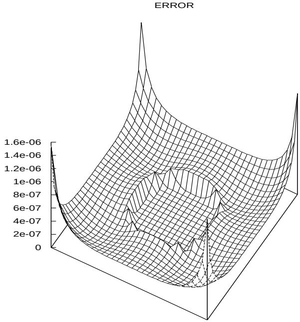


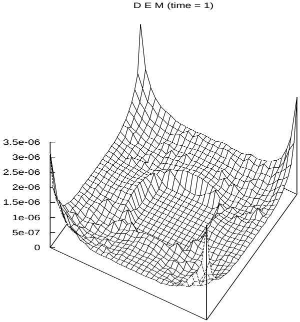
Numerical tests. Example 3

Now we consider a 2-dimensional obstacle problem with a plane obstacle and take Ω as a unit square. In the figure below, we present an approximate solution computed by the finite element method on a uniform mesh. The elements that belong to $\boldsymbol{\Omega}_\varphi$ are colored black.

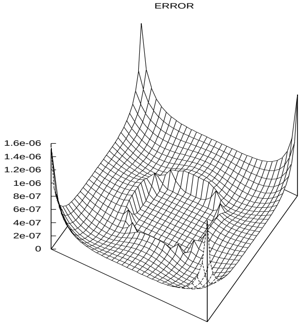


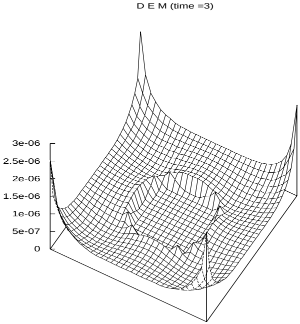
Below it is shown the distribution of local (elementwise) errors and those given by the integrand of the Majorant (for $\mathbf{t} = \mathbf{1}$).



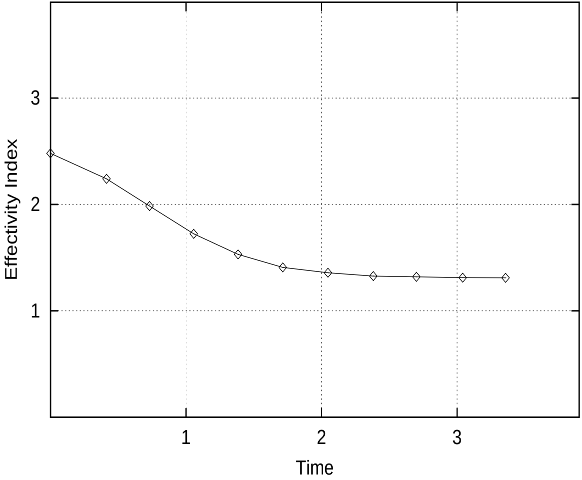


If we spend more efforts on the minimization of the Majorant ($\mathbf{t} = \mathbf{3}$), then the computed error distribution is practically the same as the true one.





Below we show the dependence of the effectivity index with respect to the CPU time used for the minimization of the Majorant



The elasto-plastic torsion problem

Let $\mathbf{\Omega}$ be a bounded domain in $\mathbb{R}^2$ with Lipschitz continuous boundary $\partial\mathbf{\Omega}$. Consider the torsion problem for a *long elasto-plastic bar* whose cross-section is the domain $\mathbf{\Omega}$. If such a bar is made of an isotropic material, then the torsion problem is reduced to the following variational inequality: find $\mathbf{u} \in \mathbf{K}$ such that

$$\int_{\mathbf{\Omega}} \nabla \mathbf{u} \cdot \nabla(\mathbf{v}-\mathbf{u})\, \mathbf{dx} \geq \mu \int_{\mathbf{\Omega}} (\mathbf{v}-\mathbf{u})\, \mathbf{dx}, \quad \forall \mathbf{v} \in \mathbf{K}, \tag{1.22}$$

where μ is a positive parameter,

$$\mathbf{K} := \{\mathbf{v} \in \mathbf{V_0} \mid |\nabla \mathbf{v}| \leq \mathbf{1} \quad \text{a.e. in } \mathbf{\Omega}\},$$

See, e.g., **G. Duvaut and J.-L. Lions**, Inequalities in mechanics and physics. Springer, Berlin-New York, 1976

Elasto–plastic torsion problem has a unique solution with a free boundary that separates the sets

$$\mathbf{\Omega_e} := \{\mathbf{x} \in \mathbf{\Omega} \mid |\nabla \mathbf{u}| < \mathbf{1}\}$$

and

$$\mathbf{\Omega_p} := \{\mathbf{x} \in \mathbf{\Omega} \mid |\nabla \mathbf{u}| = \mathbf{1}\} ,$$

which are called elastic and plastic sets, respectively.

If $\mathbf{\Omega}$ is a 1-connected domain, then $\mathbf{u}$ coincides with a solution of the following obstacle problem (see, e.g., **A. Friedman.** *Variational principles and free-boundary problems.* Wiley, NY, 1982.).

Problem. Find $\mathbf{u} \in \mathbf{K_d}$ such that

$$\mathbf{J}(\mathbf{u}) = \inf_{\mathbf{v} \in \mathbf{K_d}} \mathbf{J}(\mathbf{v}), \quad \mathbf{J}(\mathbf{v}) = \frac{1}{2} \int_{\mathbf{\Omega}} (|\nabla \mathbf{v}|^2 - \mu \mathbf{v}) \mathbf{dx},$$

where

$$\mathbf{K_d} := \{ \mathbf{v} \in \mathbf{V_0} \mid |\mathbf{v}| \leq \mathbf{d}(\mathbf{x}, \partial \mathbf{\Omega}) \text{ for a. e. } \mathbf{x} \in \mathbf{\Omega} \}$$

and $\mathbf{d}(\mathbf{x}, \partial\mathbf{\Omega})$ denotes the distance between $\mathbf{x}$ and $\partial\mathbf{\Omega}$.

It is easy to see that we arrived at a special type obstacle problem. Now, we can use the estimates (11.16) or (11.21) with

$$\varphi = -\mathbf{d}(\mathbf{x}, \partial\boldsymbol{\Omega}) \qquad \psi = \mathbf{d}(\mathbf{x}, \partial\boldsymbol{\Omega}).$$

In particular, if $\mathbf{v}$ has a fixed sign in $\boldsymbol{\Omega}$ (e.g., $\mathbf{v} \geq \mathbf{0}$), then (11.16) implies the estimate

$$\begin{aligned} \|\!| \mathbf{v} - \mathbf{u} \,|\!\|^2 \leq & \left(\mathbf{1} + \frac{\mathbf{1}}{\boldsymbol{\beta}}\right) \|\!| \nabla \mathbf{v} - \boldsymbol{\tau} \,|\!\|^2 \\ & + \mathbf{C}_{\boldsymbol{\Omega}}^2 (\mathbf{1} + \boldsymbol{\beta}) \left[\int\limits_{\boldsymbol{\Omega}_{\mathbf{e}}(\mathbf{v})} (\mathbf{div}\boldsymbol{\tau} + \mu)^2 \, \mathbf{dx} + \int\limits_{\boldsymbol{\Omega}_{\mathbf{p}}(\mathbf{v})} \langle\, \mathbf{div}\boldsymbol{\tau} + \mu \,\rangle_{\ominus}^2 \, \mathbf{dx} \right]. \end{aligned} \tag{11.23}$$

In (11.23), $\boldsymbol{\Omega}_{\mathbf{e}}(\mathbf{v})$ and $\boldsymbol{\Omega}_{\mathbf{p}}(\mathbf{v})$ are the elastic and plastic sets **defined by the approximate solution** $\mathbf{v} \in \mathbf{K}$ and $\mathbf{C}_{\boldsymbol{\Omega}}$ is a constant in the Friedrichs–Poincaré inequality.

We end up this lecture with several pictures that show some results for the elasto–plastic torsion problem.

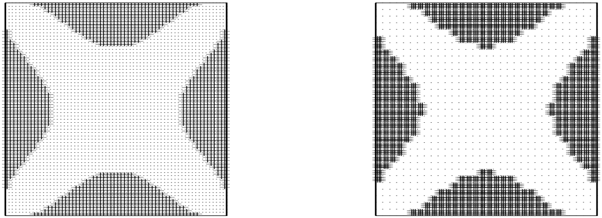
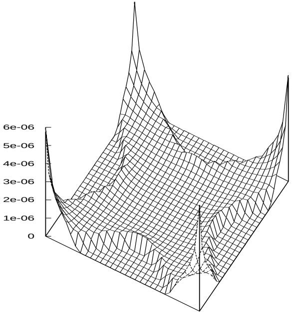
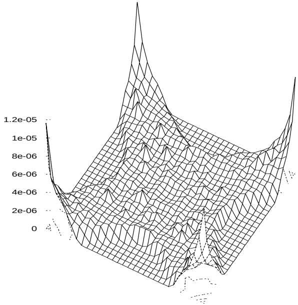


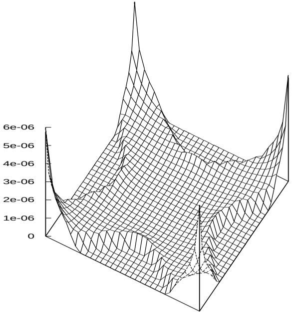
Figure: Elastic and plastic zones for $f = 5$ computed on two different meshes

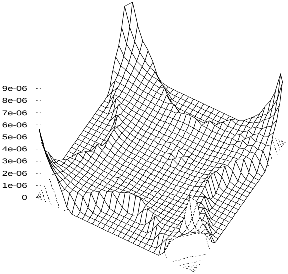
Below it is shown the distribution of local error and the distribution computed by the integrand of the Majorant for $\mathbf{t} = \mathbf{1}$.



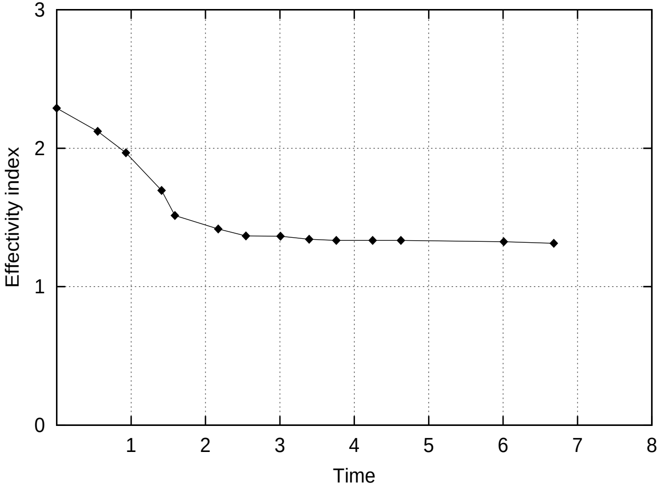


On this figure, it is shown the distribution of local error and the distribution computed by the integrand of the Majorant for $\mathbf{t = 3}$.





On this picture, we present the dependence of the effectivity index with respect to CPU time.



Lecture 12.
FUNCTIONAL A POSTERIORI ESTIMATES FOR NONLINEAR VARIATIONAL PROBLEMS

The objective of this lecture is to 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 $\mathbf{F}:\mathbf{V}\longrightarrow\mathbb{R}$ is a convex lower semicontinuous functional, $\mathbf{G}:\mathbf{Y}\longrightarrow\mathbb{R}$ is a uniformly convex functional, $\mathbf{V}$ and $\mathbf{Y}$ are reflexive Banach spaces and $\mathbf{\Lambda}:\mathbf{V}\longrightarrow\mathbf{Y}$ is a bounded linear operator.

General variational problem

Consider the general variational problem: find $\mathbf{u}$ in a Banach space V such that

$$\mathbf{J}(\mathbf{u}, \boldsymbol{\Lambda}\mathbf{u}) = \inf_{\mathbf{v}\in\mathbf{V}} \mathbf{J}(\mathbf{v}, \boldsymbol{\Lambda}\mathbf{v}), \tag{12.1}$$

where $\mathbf{J}(\mathbf{v}) = \mathbf{F}(\mathbf{v}) + \mathbf{G}(\boldsymbol{\Lambda}\mathbf{v})$, $\mathbf{F}$ is a convex, lower semicontinuous functional, $\mathbf{G}$ is a uniformly convex functional and $\boldsymbol{\Lambda} : \mathbf{V} \longrightarrow \mathbf{Y}$ 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

$$\|\boldsymbol{\Lambda}\mathbf{w}\| \geq \mathbf{c_0} \, \|\mathbf{w}\|_{\mathbf{V}} \quad \forall \, \mathbf{w} \in \mathbf{V} \,, \tag{12.2}$$

where $\mathbf{c_0}$ is a positive constant independent of $\mathbf{w}$.
In addition to $\boldsymbol{\Lambda}$, we introduce its conjugate $\boldsymbol{\Lambda}^* : \mathbf{Y}^* \longrightarrow \mathbf{V}^*$. This amounts to say that

$$\langle\langle \mathbf{y}^*, \boldsymbol{\Lambda}\mathbf{v} \rangle\rangle = \langle \boldsymbol{\Lambda}^*\mathbf{y}^*, \mathbf{v} \rangle \;\; \forall \mathbf{y}^* \in \mathbf{Y}^*, \, \mathbf{v} \in \mathbf{V} \,. \tag{12.3}$$

$\mathbf{J}(\mathbf{v}, \boldsymbol{\Lambda}\mathbf{v}) := \mathbf{F}(\mathbf{v}) + \mathbf{G}(\boldsymbol{\Lambda}\mathbf{v})$. is assumed to be coercive on $\mathbf{V}$, i.e.

$$\mathbf{J}(\mathbf{v}, \boldsymbol{\Lambda}\mathbf{v}) \longrightarrow +\infty \quad \text{if} \quad \|\mathbf{v}\|_{\mathbf{V}} \longrightarrow +\infty.$$

Primal and Dual Problems

Problem $\mathcal{P}$. *Find* $\mathbf{u} \in \mathbf{V}$ *such that*

$$\mathbf{J}(\mathbf{u}, \mathbf{\Lambda u}) = \mathbf{inf}\,\mathcal{P} := \inf_{\mathbf{v} \in \mathbf{V}} \mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}). \tag{12.4}$$

The problem dual to (12.4 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*

$$-\mathbf{J}^*(\mathbf{\Lambda}^*\mathbf{p}^*, -\mathbf{p}^*) = \mathbf{sup}\,\mathcal{P}^* := \sup_{\mathbf{y}^* \in \mathbf{Y}^*} -\mathbf{J}^*(\mathbf{\Lambda}^*\mathbf{y}^*, -\mathbf{y}^*), \tag{12.5}$$

$$\mathbf{J}^*(\mathbf{\Lambda}^*\mathbf{y}^*, -\mathbf{y}^*) := \mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) + \mathbf{G}^*(-\mathbf{y}^*),$$

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* $\boldsymbol{\Lambda}\mathbf{u_0} \in \mathbf{Y}$*, then there exists a minimizer* $\mathbf{u}$ *to Problem* $\mathcal{P}$ *and a maximizer* $\mathbf{p}^*$ *to Problem* $\mathcal{P}^*$*. Besides,*

$$\inf \mathcal{P} = \sup \mathcal{P}^* \tag{12.6}$$

and the following duality relations hold

$$\begin{aligned} &\textbf{(i)} \quad \mathbf{F}(\mathbf{u}) + \mathbf{F}^*(\boldsymbol{\Lambda}^*\mathbf{p}^*) - \langle \boldsymbol{\Lambda}^*\mathbf{p}^*, \mathbf{u} \rangle = \mathbf{0}, \\ &\textbf{(ii)} \quad \mathbf{G}(\boldsymbol{\Lambda}\mathbf{u}) + \mathbf{G}^*(-\mathbf{p}^*) + \langle\langle \mathbf{p}^*, \boldsymbol{\Lambda}\mathbf{u} \rangle\rangle = \mathbf{0}. \end{aligned} \tag{12.7}$$

Above relations are equivalent to

$$\textbf{(i)}\ \boldsymbol{\Lambda}^*\mathbf{p}^* \in \partial\mathbf{F}(\mathbf{u}), \qquad \textbf{(ii)}\ -\mathbf{p}^* \in \partial\mathbf{G}(\boldsymbol{\Lambda}\mathbf{u}).$$

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 $\mathbf{B}(\mathbf{0}, \delta) := \{\mathbf{y} \in \mathbf{Y} \mid \|\mathbf{y}\| < \delta\}$ 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 $\mathbf{y} = \mathbf{O}_\mathbf{y}$ is and

$$\mathbf{G}\left(\frac{\mathbf{y_1}+\mathbf{y_2}}{\mathbf{2}}\right) + \mathbf{\Phi}_\delta(\mathbf{y_2} - \mathbf{y_1}) \leq \frac{\mathbf{1}}{\mathbf{2}}\left(\mathbf{G}(\mathbf{y_1}) + \mathbf{G}(\mathbf{y_2})\right) \quad \forall\, \mathbf{y_1}, \mathbf{y_2} \in \mathbf{B}(\mathbf{0}, \delta).$$

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 $\mathbf{y}^* \in \mathbf{Y}^*$

$$\mathbf{\Phi}_\delta\left(\mathbf{\Lambda}(\mathbf{v}-\mathbf{u})\right) \leq \mathbf{M}_\oplus(\mathbf{v},\mathbf{y}^*) := \mathbf{D_F}(\mathbf{\Lambda}^*\mathbf{y}^*,\mathbf{v}) + \mathbf{D_G}(\mathbf{y}^*,\mathbf{\Lambda v}) \tag{12.8}$$

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

$$\boldsymbol{\Phi}_\delta\left(\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\right)+\mathbf{G}(\boldsymbol{\Lambda}(\tfrac{\mathbf{v}+\mathbf{u}}{2}))+\mathbf{F}(\tfrac{\mathbf{v}+\mathbf{u}}{2})\leq \\ \tfrac{1}{2}\Big[(\mathbf{F}(\mathbf{v})+\mathbf{G}(\boldsymbol{\Lambda}\mathbf{v}))+(\mathbf{F}(\mathbf{u})+\mathbf{G}(\boldsymbol{\Lambda}\mathbf{u}))\Big].$$

The element $\mathbf{u}$ is a minimizer, therefore

$$\mathbf{G}(\boldsymbol{\Lambda}\mathbf{u})+\mathbf{F}(\mathbf{u})=\mathbf{J}(\mathbf{u})\;\leq\;\mathbf{G}(\boldsymbol{\Lambda}\left(\tfrac{\mathbf{u}+\mathbf{v}}{2}\right))+\mathbf{F}(\tfrac{\mathbf{u}+\mathbf{v}}{2})$$

and we have

$$\boldsymbol{\Phi}_\delta\left(\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\right)+\mathbf{G}(\boldsymbol{\Lambda}\mathbf{u})+\mathbf{F}(\mathbf{u})\leq \\ \tfrac{1}{2}\Big[(\mathbf{F}(\mathbf{v})+\mathbf{G}(\boldsymbol{\Lambda}\mathbf{v}))+(\mathbf{F}(\mathbf{u})+\mathbf{G}(\boldsymbol{\Lambda}\mathbf{u}))\Big].$$

From the above we observe that

$$\mathbf{\Phi}_\delta\,(\mathbf{\Lambda e}) \leq \frac{1}{2}\Big[(\mathbf{F}(\mathbf{v}) + \mathbf{G}(\mathbf{\Lambda v})) - (\mathbf{F}(\mathbf{u}) + \mathbf{G}(\mathbf{\Lambda u}))\Big] =$$
$$= \frac{1}{2}(\mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}) - \mathbf{J}(\mathbf{u}, \mathbf{\Lambda u})) \quad \forall \mathbf{v} \in \mathbf{B}(\mathbf{0}, \delta)\,.$$

In view of Theorem 1,

$$\mathbf{J}(\mathbf{u}, \mathbf{\Lambda u}) = \mathbf{inf}\,\mathcal{P} = \mathbf{sup}\,\mathcal{P}^* = -\mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{p}^*) - \mathbf{G}^*(-\mathbf{p}^*).$$

Since $\mathbf{p}^*$ is a solution of the dual problem, we have

$$-\mathbf{J}^*(\mathbf{\Lambda}^*\mathbf{p}^*, -\mathbf{p}^*) \geq -\mathbf{J}^*(\mathbf{\Lambda}^*\mathbf{y}^*, -\mathbf{y}^*) \quad \forall \mathbf{y}^* \in \mathbf{Y}^*,$$

so that

$$\mathbf{J}(\mathbf{u}, \mathbf{\Lambda u}) \geq -\mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) - \mathbf{G}^*(-\mathbf{y}^*).$$

Therefore

$$\begin{aligned}\boldsymbol{\Phi}_\delta\left(\boldsymbol{\Lambda}\mathbf{e}\right) &\leq \frac{1}{2}\left(\mathbf{F}(\mathbf{v})+\mathbf{G}(\boldsymbol{\Lambda}\mathbf{v})+\mathbf{F}^*(\boldsymbol{\Lambda}^*\mathbf{p}^*)+\mathbf{G}^*(-\mathbf{p}^*)\right) \leq \\ &\leq \frac{1}{2}\left(\mathbf{F}(\mathbf{v})+\mathbf{G}(\boldsymbol{\Lambda}\mathbf{v})+\mathbf{F}^*(\boldsymbol{\Lambda}^*\mathbf{y}^*)+\mathbf{G}^*(-\mathbf{y}^*)\right)\,.\end{aligned}$$

However, by (12.3) we observe that

$$\langle\langle \mathbf{y}^*, \boldsymbol{\Lambda}\mathbf{v}\rangle\rangle - \langle \boldsymbol{\Lambda}^*\mathbf{y}^*, \mathbf{v}\rangle = \mathbf{0}\ \ \forall \mathbf{y}^* \in \mathbf{Y}^*,\ \mathbf{v} \in \mathbf{V}\,.$$

We add this zero term to the above relation and obtain the required estimate.

$\square$

Comments

The right–hand side of (12.8 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 (12.7)(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

$$\mathbf{\Phi}_\delta\,(\mathbf{\Lambda e}) + \varphi_\delta(\mathbf{e}) \quad \leq \ \tfrac{1}{2}(\mathbf{J}(\mathbf{v}, \mathbf{\Lambda v}) - \mathbf{J}(\mathbf{u}, \mathbf{\Lambda u})) \qquad (12.9)$$

and, as a result, (12.8) is replaced by the strengthened estimate

$$\mathbf{\Phi}_\delta\,(\mathbf{\Lambda e}) + \varphi_\delta(\mathbf{e}) \ \leq \ \mathbf{M}_\oplus\,(\mathbf{v}, \mathbf{y}^*) \quad \forall \mathbf{y}^* \in \mathbf{Y}^*\,. \qquad (12.10)$$

It is not difficult to verify that

$$
\begin{aligned}
&\mathbf{M}_{\oplus}(\mathbf{v},\mathbf{y}^*) - \mathbf{M}_{\oplus}(\mathbf{v},\mathbf{p}^*) = \\
&= \tfrac{1}{2}\left(\mathbf{F}(\mathbf{v}) + \mathbf{F}^*(\boldsymbol{\Lambda}^*\mathbf{y}^*) - \langle \boldsymbol{\Lambda}^*\mathbf{y}^*, \mathbf{v}\rangle + \mathbf{G}(\boldsymbol{\Lambda}\mathbf{v}) + \mathbf{G}^*(-\mathbf{y}^*) + \langle\!\langle \mathbf{y}^*, \boldsymbol{\Lambda}\mathbf{v}\rangle\!\rangle\right) - \\
&\tfrac{1}{2}\left(\mathbf{F}(\mathbf{v}) + \mathbf{F}^*(\boldsymbol{\Lambda}^*\mathbf{p}^*) - \langle \boldsymbol{\Lambda}^*\mathbf{p}^*, \mathbf{v}\rangle + \mathbf{G}(\boldsymbol{\Lambda}\mathbf{v}) + \mathbf{G}^*(-\mathbf{p}^*) + \langle\!\langle \mathbf{p}^*, \boldsymbol{\Lambda}\mathbf{v}\rangle\!\rangle\right) = \\
&\qquad = \mathbf{J}^*(\boldsymbol{\Lambda}^*\mathbf{y}^*, -\mathbf{y}^*) - \mathbf{J}^*(\boldsymbol{\Lambda}^*\mathbf{p}^*, -\mathbf{p}^*) = \geq \mathbf{0}.
\end{aligned}
$$

Therefore, for any $\mathbf{v}$ the right-hand side of (12.8) 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

$$
-\boldsymbol{\sigma}^*(\mathbf{v}) \in \partial \mathbf{G}(\boldsymbol{\Lambda}\mathbf{v}) \subset \mathbf{Y}^*.
$$

In this case,

$$\mathbf{M}_{\mathbf{G}}(\boldsymbol{\sigma}^*(\mathbf{v}), \boldsymbol{\Lambda}\mathbf{v}) = \mathbf{0}$$

and we get the estimate

$$\boldsymbol{\Phi}_\delta\left(\boldsymbol{\Lambda}\mathbf{e}\right) \,\leq\, \mathbf{M}_{\mathbf{F}}(\boldsymbol{\Lambda}^*\boldsymbol{\sigma}^*(\mathbf{v}), \mathbf{v}) \qquad (12.11)$$

whose right–hand side depends on $\mathbf{v}$ only.
However, the estimate (12.11) cannot be directly applied in one practically important case which we consider below.

Problems with linear functional F

Let

$$\mathbf{F}(\mathbf{v}) = \langle \boldsymbol{\ell}^*, \mathbf{v} \rangle, \quad \boldsymbol{\ell}^* \in \mathbf{V}^*. \tag{12.12}$$

Since

$$\mathbf{F}^*(\mathbf{v}^*) = \sup_{\mathbf{v}\in\mathbf{V}} \langle \mathbf{v}^* - \boldsymbol{\ell}^*, \mathbf{v} \rangle = \begin{cases} 0 & \text{if } \mathbf{v}^* = \boldsymbol{\ell}^*, \\ +\infty & \text{if } \mathbf{v}^* \neq \boldsymbol{\ell}^* \end{cases}$$

we see that

$$\begin{aligned} \mathbf{M}_{\mathbf{F}} &= \langle \boldsymbol{\ell}^*, \mathbf{v} \rangle + \mathbf{F}^*(\boldsymbol{\Lambda}^* \mathbf{y}^*) - \langle \boldsymbol{\Lambda}^* \mathbf{y}^*, \mathbf{v} \rangle = \\ &= \langle \boldsymbol{\ell}^* - \boldsymbol{\Lambda}^* \mathbf{y}^*, \mathbf{v} \rangle + \mathbf{F}^*(\boldsymbol{\Lambda}^* \mathbf{y}^*). \end{aligned}$$

$$\mathbf{M_F}(\mathbf{\Lambda}^*\mathbf{y}^*, \mathbf{v}) = \mathbf{F}^*(\mathbf{\Lambda}^*\mathbf{y}^*) + \langle \ell^* - \mathbf{\Lambda}^*\mathbf{y}^*, \mathbf{v} \rangle = \begin{cases} 0 & \text{if } \mathbf{y}^* \in \mathbf{Q}^*_{\boldsymbol{\ell}}, \\ +\infty & \text{if } \mathbf{y}^* \notin \mathbf{Q}^*_{\boldsymbol{\ell}}, \end{cases}$$

where

$$\mathbf{Q}^*_{\boldsymbol{\ell}} := \{\mathbf{y}^* \in \mathbf{Y}^* \mid \langle \mathbf{\Lambda}^*\mathbf{y}^*, \mathbf{w} \rangle = \langle \ell^*, \mathbf{w} \rangle \ \forall \mathbf{w} \in \mathbf{V}\} \ .$$

In general, above defined $\boldsymbol{\sigma}^*$ does not belong to $\mathbf{Q}^*_{\boldsymbol{\ell}}$, so that the right hand side of (12.11) can become infinite. Therefore, the aim of our subsequent analysis is to obtain a modified error majorant $\widetilde{\mathbf{M}}_{\oplus}(\mathbf{v}, \mathbf{y}^*)$ which is finite for all $\mathbf{v} \in \mathbf{V}$ and all $\mathbf{y}^* \in \mathbf{Y}^*$.

Let $\mathbf{\Pi}(\mathbf{y}) \geq \mathbf{0}$ for all $\mathbf{y}$ $\mathbf{\Pi}(\mathbf{0}) = \mathbf{0}$. By $\mathbf{\Pi}^* : \mathbf{Y}^* \longrightarrow \mathbb{R}_+$ we denote the functional conjugate of $\mathbf{\Pi}$. For this pair the Joung–Fenchel inequality

$$\langle\langle \xi^*, \xi \rangle\rangle \leq \mathbf{\Pi}^*(\xi^*) + \mathbf{\Pi}(\xi) \quad \forall \xi \in \mathbf{Y}, \;\; \xi^* \in \mathbf{Y}^*$$

holds.

For the sake of simplicity, we set

$$\mathbf{\Pi}(\mathbf{y}) = \pi(\|\mathbf{y}\|) \quad \text{and} \quad \mathbf{\Pi}^*(\mathbf{y}^*) = \pi^*(\|\mathbf{y}^*\|_*).$$

General form of the Deviation Majorant

$$\boldsymbol{\Phi}_\delta\left(\boldsymbol{\Lambda}(\mathbf{v}-\mathbf{u})\right) \leq \mathbf{M}_{\mathbf{D}}(\mathbf{y}^*, \boldsymbol{\Lambda}\mathbf{v}) + \mathbf{M}_{\mathbf{R}}(\mathbf{y}^*)\,, \tag{12.13}$$

where

$$\mathbf{M}_{\mathbf{D}}(\mathbf{y}^*, \boldsymbol{\Lambda}\mathbf{v}) = \mathbf{D}_{\mathbf{G}}(\mathbf{y}^*, \boldsymbol{\Lambda}\mathbf{v}) + \tfrac{1}{2}\pi\left(\|\mathbf{G}^{*\prime}(-\mathbf{y}^*) - \boldsymbol{\Lambda}\mathbf{v}\|\right), \tag{12.14}$$

$$\mathbf{M}_{\mathbf{R}}(\mathbf{y}^*) = \inf_{\mathbf{q}^* \in \mathbf{Q}^*_\ell} \pi^*(\|\mathbf{q}^* - \mathbf{y}^*\|_*). \tag{12.15}$$

Examples

Set $\mathbf{\Lambda v} := \nabla \mathbf{v}$ and consider variational problems for the functional

$$\mathbf{J}(\mathbf{v}, \nabla \mathbf{v}) = \int_{\mathbf{\Omega}} (\mathbf{g}(\nabla \mathbf{v}) + \mathbf{f}(\mathbf{v}))\, \mathbf{dx}\,.$$

Now $\mathbf{G}$ and $\mathbf{F}$ are integral functionals whose integrands $\mathbf{g} : \mathbb{R}^{\mathbf{d}} \longrightarrow \mathbb{R}$ and $\mathbf{f} : \mathbb{R} \longrightarrow \mathbb{R}$ are convex differentiable functions. Denote their conjugate functions $\mathbf{g}^*$ and $\mathbf{f}^*$, respectively. The spaces $\mathbf{Y}$ and $\mathbf{Y}^*$ we identify with the Lebesque spaces $\mathbf{L}^{\boldsymbol{\alpha}}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}})$ and $\mathbf{L}^{\boldsymbol{\alpha}_*}(\mathbf{\Omega}, \mathbb{R}^{\mathbf{d}})$, where $\boldsymbol{\alpha}_* = \frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}-\mathbf{1}}$, $\boldsymbol{\alpha} > 1$ is taken such that the above integral has sense. In the considered case,

$$\langle \mathbf{y}^*, \mathbf{y} \rangle := \int_{\mathbf{\Omega}} \mathbf{y}^* \cdot \mathbf{y}\, \mathbf{dx} \ \text{ and } \ \mathbf{\Lambda}^* \mathbf{y}^* := -\mathbf{div} \mathbf{y}^* \in \mathbf{V}^*.$$

Example 1.

Let $\mathbf{g}(\mathbf{y}) = \frac{1}{2}\mathbf{A}\mathbf{y}\cdot\mathbf{y}$, where $\mathbf{A}$ is a symmetric real matrix satisfying the conditions

$$\nu_1 \,|\boldsymbol{\eta}|^2 \leq \mathbf{A}\boldsymbol{\eta}\cdot\boldsymbol{\eta} \leq \nu_2\,|\boldsymbol{\eta}|^2 \quad \forall\boldsymbol{\eta}\in\mathbb{R}^{\mathbf{d}},$$

for some $\nu_2 \geq \nu_1 > 0$. It is straightforward to check that the functional $\mathbf{G}$ is uniformly convex on any ball. The two parts of the error majorant $\mathbf{M}_\oplus$ (cf. (12.8) are given by the relations

$$\mathbf{D_G}(\mathbf{y}^*,\boldsymbol{\Lambda}\mathbf{v}) = \frac{1}{4}\int_\Omega \left(\mathbf{A}\nabla\mathbf{v}\cdot\nabla\mathbf{v} + \mathbf{A}^{-1}\mathbf{y}^*\cdot\mathbf{y}^* - \nabla\mathbf{v}\cdot\mathbf{y}^*\right)\mathbf{dx},$$

$$\mathbf{D_F}(\boldsymbol{\Lambda}^*\mathbf{y}^*,\mathbf{v}) = \frac{1}{2}\int_\Omega (\mathbf{f}(\mathbf{v}) - \mathbf{y}^*\cdot\nabla\mathbf{v})\mathbf{dx} + \frac{1}{2}\sup_{\mathbf{w}\in\mathbf{V}}\int_\Omega (\mathbf{y}^*\cdot\nabla\mathbf{w} - \mathbf{f}(\mathbf{w}))\mathbf{dx}.$$

If the function $\mathbf{f}^*(-\mathbf{div}\mathbf{y}^*)$ is summable then we arrive at a more symmetric expression

$$\mathbf{D_F}(\mathbf{v}, \mathbf{y}^*) \leq \frac{1}{2} \int_{\Omega} \left(\mathbf{f}(\mathbf{v}) + \mathbf{f}^*(-\mathbf{div}\mathbf{y}^*) - \mathbf{y}^* \cdot \nabla \mathbf{v}\right) \mathbf{dx}\,.$$

In particular, if $\mathbf{f}(\mathbf{v}) = \frac{\lambda}{2}\mathbf{v}^2 + \mu\mathbf{v}$, where $\mu \in \mathbb{R}$ and $\lambda \in \mathbb{R}_+$, then

$$\mathbf{f}^*(\mathbf{v}^*) = \frac{1}{2\lambda}(\mathbf{v}^* - \mu)^2.$$

We note that this case is related to the equation

$$\mathbf{divA}\nabla\mathbf{u} - \lambda\mathbf{u} + \mu = \mathbf{0}.$$

In this case, $\alpha = 2$ and for any

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

we obtain

$$\mathbf{M_F}(\mathbf{v}, \mathbf{y}^*) \leq \frac{\mathbf{1}}{\mathbf{4}\boldsymbol{\lambda}} \|\boldsymbol{\lambda}\mathbf{v} + \mathbf{div}\mathbf{y}^* + \mu\|^{\mathbf{2}}_{\boldsymbol{\Omega}} \ ,$$

Both functionals **G** and **F** are uniformly convex and we can take

$$\boldsymbol{\Phi}(\nabla \mathbf{e}) = \frac{\mathbf{1}}{\mathbf{4}} \int_{\boldsymbol{\Omega}} \mathbf{A}\nabla \mathbf{e} \cdot \nabla \mathbf{e} \, \mathbf{dx},$$

$$\varphi(\mathbf{e}) = \frac{\boldsymbol{\lambda}}{\mathbf{4}} \int_{\boldsymbol{\Omega}} |\mathbf{e}|^{\mathbf{2}} \, \mathbf{dx}.$$

Thus, we arrive at the estimate of deviation in the following form:

$$\int_\Omega \mathbf{A}\nabla(\mathbf{v}-\mathbf{u})\cdot\nabla(\mathbf{v}-\mathbf{u})\,\mathbf{dx} + \boldsymbol{\lambda}\,\|\mathbf{v}-\mathbf{u}\|^2_\Omega \le$$

$$\le \int_\Omega (\mathbf{A}^{-1}\mathbf{y}^* + \nabla\mathbf{v})\cdot(\mathbf{y}^* + \mathbf{A}\nabla\mathbf{v})\,\mathbf{dx} +$$

$$+\frac{1}{\boldsymbol{\lambda}}\,\|\boldsymbol{\lambda}\mathbf{v} + \mathbf{div}\mathbf{y}^* + \boldsymbol{\mu}\|^2_\Omega\,.$$

Example 2

Consider the problem with

$$\mathbf{G}(\mathbf{y}) = \tfrac{1}{2}(\mathcal{A}\mathbf{y}, \mathbf{y}) + \mathbf{\Psi}(\mathbf{y}), \quad \mathbf{F}(\mathbf{v}) = < \ell, \mathbf{v} >,$$

where $\mathbf{\Psi} : \mathbf{Y} \to \mathbb{R}$ is a convex continuous functional. Note that if $\Psi \equiv 0$, then

$$\begin{aligned} \mathbf{D_G}(\mathbf{\Lambda v}, \mathbf{p}^*) = & \ \tfrac{1}{2}(\mathcal{A}\mathbf{\Lambda v}, \mathbf{\Lambda v}) + \tfrac{1}{2}(\mathcal{A}\mathbf{\Lambda u}, \mathbf{\Lambda u}) - (\mathbf{\Lambda v}, \mathcal{A}\mathbf{\Lambda u}) \\ & = \tfrac{1}{2} \,|\!|\!|\, \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \,|\!|\!|^{\mathbf{2}} . \end{aligned}$$

We will measure the error in terms of the above norm generated by the operator $\mathbf{A}$.

In this case, the deviation estimate is as follows:

$$\frac{1}{2} \parallel \mathbf{\Lambda}(\mathbf{v} - \mathbf{u}) \parallel^2 \leq (1 + \beta)\mathbf{D_G}(\mathbf{\Lambda v}, \mathbf{y}^*) +$$

$$+ \left(1 + \frac{1}{\beta}\right) \mathbf{C}_{\mathbf{\Omega}}^2 \|\mathbf{\Lambda}^* \mathbf{y}^* + \ell\|^2,$$

where $\mathbf{C}_{\mathbf{\Omega}}$ depends on Ω and $\mathbf{A}$.

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.

We end up this lecture course with concise exposition of two important problems closely related with functional type a posteriori estimates.

- Evaluation of errors in terms of local quantities;
- Evaluation of modeling errors.

Indication of local errors

Integrand of the Majorant is a good error indicator.

$$\mathbf{M}_{\oplus} = \int_{\mathbf{\Omega}} \mu(\mathbf{x})\mathbf{dx}.$$

It is proved that if $\mathbf{M}_{\oplus} \longrightarrow \| \mathbf{u} - \mathbf{v} \|$, then

$$\mu(\mathbf{x}) \rightarrow \mathbf{e}(\mathbf{x}) := |\nabla(\mathbf{u} - \mathbf{v})|(\mathbf{x}) \quad \text{in the sense of measures}$$

This means that for any $\delta > 0$

$$\mathbf{measE}_{\delta} \rightarrow \mathbf{0}$$

where $\mathbf{E}_{\delta} := \{\mathbf{x} \in \mathbf{\Omega} \,|\, |\mu(\mathbf{x}) - \mathbf{e}(\mathbf{x})| > \delta\}$.

Guaranteed upper bounds of local errors

GENERAL PRINCIPLE: Guaranteed upper bounds for

$$\boldsymbol{\Phi}(\mathbf{u}-\mathbf{v}) = \|\mathbf{u}-\mathbf{v}\|_\omega \ \text{ and } \ \boldsymbol{\Phi}(\mathbf{u}-\mathbf{v}) = (\boldsymbol{\ell}, \mathbf{u}-\mathbf{v}), \quad \boldsymbol{\ell} \in \mathbf{V}^*$$

are obtained by projection of the functional a posteriori estimate onto a certain subspace.

S. Repin. A posteriori estimates in local norms. J. Math. Sci. (N. Y.) 124 (2004), no. 3, 5026–5035.

S. Repin. Local a posteriori estimates for the Stokes problem. Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 318 (2004), 35, 233–245, 312–313.

Example. Local estimate for diffusion problem

Let $\omega \subset \mathbf{\Omega}$ and $\mathbf{V}_\omega := \{\mathbf{H}^1_0(\mathbf{\Omega}) \,|\, \mathbf{v} = \mathbf{const} \text{ in } \omega\}$. An upper bound of the local error is given by the estimate

$$\|\nabla(\mathbf{u}-\mathbf{v})\|^2_\omega \leq \mathbf{M}_{\oplus\omega} :=$$
$$:= \inf_{\mathbf{w}\in\mathbf{V}_\omega} \left\{(1+\beta)\|\nabla(\mathbf{v}-\mathbf{w}) - \mathbf{y}\|^2 + \frac{1+\beta}{\beta}\mathbf{C}^2_{\mathbf{\Omega}}\|\mathbf{div}\mathbf{y} + \mathbf{f}\|^2\right\}. \tag{12.16}$$

Here $\beta > 0$ and $\mathbf{y} \in \mathbf{H}(\mathbf{div}, \mathbf{\Omega})$.

Quality of the estimate

$$\| \nabla(\mathbf{u}-\mathbf{v}) \|_\omega^2 \leq \mathbf{M}_\omega^\oplus(\mathbf{v}) \leq \| \nabla(\mathbf{u}-\mathbf{v}) \|_\omega^2 + \mathbf{I}_\omega(\mathbf{v}),$$

where

$$\mathbf{I}_\omega(\mathbf{v}) := \inf_{\phi \in \mathbf{V}_\omega} \| \nabla(\mathbf{u}-\mathbf{v}-\phi) \|_{\Omega \setminus \omega}^2$$

If $\mathbf{v}-\mathbf{u} = \mu = \mathbf{const}$ on $\partial\omega$, then the function

$$\bar{\phi} := \begin{cases} u - v & \text{in } \Omega \setminus \omega; \\ \mu & \text{in } \omega \end{cases}$$

belongs to $\mathbf{V}_\omega$ and, therefore, $\mathbf{I}_\omega(\mathbf{v}) = \mathbf{0}$.

Errors in terms of goal oriented quantities

In the two above cited papers a guaranteed upper bounds for goal–oriented errors were also derived.
Basic observation

$$| \langle \boldsymbol{\ell}, \mathbf{u} - \mathbf{v} \rangle | = | \langle \boldsymbol{\ell}, \mathbf{u} - \mathbf{v} + \varphi \rangle | \qquad \forall \varphi \in \mathbf{V}_{\mathbf{0}\boldsymbol{\ell}}(\boldsymbol{\Omega}),$$

where $\mathbf{V}_{\mathbf{0}\boldsymbol{\ell}}(\boldsymbol{\Omega}) := \{\varphi \in \mathbf{H}^{\mathbf{1}}_{\mathbf{0}}(\boldsymbol{\Omega}) \mid \langle \boldsymbol{\ell}, \varphi \rangle = \mathbf{0}\}$. Therefore,

$$| \langle \boldsymbol{\ell}, \mathbf{u} - \mathbf{v} \rangle | \leq \|\boldsymbol{\ell}\| \inf_{\varphi \in \mathbf{V}_{\mathbf{0}\boldsymbol{\ell}}} \|\mathbf{u} - \mathbf{v} + \varphi\|.$$

Since $\mathbf{v} - \varphi$ can be viewed as a certain approximation, we apply the functional error estimate to the left hand side and obtain a guaranteed bound for the goal–oriented quantity.

Modeling errors

Since there are no ”absolutely exact” mathematical models, modeling errors always exist in real life mathematical modeling.
How to estimate their influence?

Let us shortly consider this question in connection with one type of modeling errors that arise in **dimension reduction models.**

$$\mathbf{\Omega} = \widehat{\mathbf{\Omega}} \times (-\mathbf{d}, +\mathbf{d}),$$

$\widehat{\mathbf{\Omega}} \in \mathbf{R}^2$ with boundary γ,

$$\mathbf{d} \ll \mathrm{diam}(\widehat{\mathbf{\Omega}}) := \sup_{(\mathbf{x}_1,\mathbf{x}_2)\in\widehat{\mathbf{\Omega}}} |\mathbf{x}_1 - \mathbf{x}_2| \, .$$

A more detailed exposition can be found in

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), no. 4, 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, No. 1, 714–715 (2004).

S. Repin, S. Sauter and A. Smolianski. A posteriori estimation of dimension reduction errors. In *Proc. 5th European Conference on Numerical Mathematics and applications, Prague 2004*, 717–725.

S. Repin. Estimates for errors in two-dimensional models of elasticity theory. *J. Math. Sci. (New York)*, 106 (2001), 3, 3027–3041.

Key idea

We can consider a solution of a $\mathbf{d-1}$ –dimensional model as an approximate solution of the $\mathbf{d}$–dimensional one. Since deviation estimates are valid for all conforming approximations in the energy space, we may somehow project $\mathbf{d-1}$–dimensional solution to the energy space of the $\mathbf{d}$–dimensional problem and use the Deviation Estimate for the estimation of the respective error.

Example. Plain stress problem as a model of 3D linear elasticity one

Here, 3D solution $(\mathbf{u}, \boldsymbol{\sigma})$ is approximated by the 2D one $(\widehat{\mathbf{u}}, \widehat{\boldsymbol{\sigma}})$, where $\widehat{\mathbf{u}} = (\widehat{\mathbf{u}}_1, \widehat{\mathbf{u}}_2)$ and $\widehat{\boldsymbol{\sigma}}$ is a $\mathbf{2} \times \mathbf{2}$ tensor. Let

$$\widetilde{\mathbf{u}} = (\widehat{\mathbf{u}}_{\mathbf{1}}, \widehat{\mathbf{u}}_{\mathbf{2}}, \phi(\mathbf{x}_{\mathbf{1}}, \mathbf{x}_{\mathbf{2}}, \mathbf{x}_{\mathbf{3}})); \; \widetilde{\boldsymbol{\sigma}}_{\alpha\beta} = \widehat{\boldsymbol{\sigma}}_{\alpha\beta}, \quad \widetilde{\boldsymbol{\sigma}}_{\mathbf{3}\alpha} = \mathbf{0},$$

where $\varphi \in \mathbf{H}^{\mathbf{1}}(\boldsymbol{\Omega})$ and meets boundary conditions $(\mathbf{u_{03}})$ on the Dirichlet part of $\partial\boldsymbol{\Omega}$. Then, we have

An estimate of the dimension reduction error

$$\mathbf{C}_\varepsilon \parallel \varepsilon(\widetilde{\mathbf{u}} - \mathbf{u})\|^2_\Omega + \mathbf{C}_\tau \parallel \widetilde{\boldsymbol{\sigma}} - \boldsymbol{\sigma}\|^2_\Omega \leq$$

$$\leq \left(\frac{\mathbf{K_0}}{2} + \frac{2\mu}{3}\right) \int_\Omega \left(\rho(\widehat{\mathbf{u}}_{1,1} + \widehat{\mathbf{u}}_{2,2}) + \varphi_{,3}\right)^2 \mathbf{dx} + \frac{\mu}{2} \int_\Omega \left(\varphi^2_{,1} + \varphi^2_{,2}\right) \mathbf{dx}$$

See the proof in
S. Repin. *J. Math. Sci. New York*, v. 106,3, 2001.
Here $\boldsymbol{\mu}$ and $\mathbf{K_0}$ are elasticity coefficients $\rho = \frac{3\mathbf{K_0} - 2\mu}{3\mathbf{K_0} + 4\boldsymbol{\mu}}$,,

$$\mathbf{C}_\varepsilon = \min\{\frac{1}{\boldsymbol{\mu}}, \frac{2}{3\boldsymbol{\mu}}\},\ \mathbf{C}_\tau = \min\{4\boldsymbol{\mu}, 6\mathbf{K_0}\}.$$

General diffusion type equations. Non-plane domains

In
S. Repin, S. Sauter, A. Smolianski. *SIAM J. Numer. Anal.* 2004.
computable estimates for dimension reduction models we derived for general diffusion type problem

$$\mathbf{div}(\mathbf{A}\nabla\mathbf{u}) + \mathbf{f} = \mathbf{0}, \quad \mathbf{u} = \mathbf{u_0}\ \partial\boldsymbol{\Omega}$$

for "thin" domains of the form $\boldsymbol{\Omega} = \boldsymbol{\Gamma} \times [-\mathbf{d}, \mathbf{d}]$, where Γ is a certain surface in 3D.

Below is the list of publications related to the topics discussed in the Lectures. In brackets, it shown the number of a lecture related to a particular publication.

- Y. Achdou, C. Bernardi and F. Coquel, A priori and a posteriori analysis of finite volume discretizations of Darcy's equations. Numer. Math. **96**(1) (2003), 17–42.
- M. Ainsworth and J. T. Oden. A posteriori error estimation in the finite element method, *Numer. Math.*, 60(1992) 429-463. (L2)
- M. Ainsworth and J. T. Oden. A unified approach to a posteriori error estimation using element residual methods, *Numer. Math.*, 65(1993) 23-50. (L2)
- M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Wiley and Sons, New York, 2000. (L2)
- 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. (L2)
- 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. (L2)

- A. Alonso, Error estimators for a mixed method. Numer. Math. **74** (1996), 385–395.
- 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. (L2)
- 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). (L11)
- A. M. Arthurth. Complementary variational principles, Clarendon Press, Oxford, 1980. (L2)
- G. Auchmuty. A posteriori error estimates for linear equations, *Numer. Math.*, 61(1992), 1-6. (L2)
- O. Axelsson. *Iterative solution methods*. Cambridge University Press, Cambridge, 1994. (L10)

- 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. (L1,2)
- I. Babuška. The finite element method with Lagrangian multipliers, *Numer. Math.*, 20(1973), 179-192. (L7)
- 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. (L2)
- 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. (L2, a posteriori estimate for pollution errors, local estimates)
- I. Babuška and J. E. Osborn. Can a finite element method perform arbitrarily badly?, *Math. Comput.*, 69(2000), 230, 443-462. (L1)

- I. Babuška and W. C. Rheinboldt. A-posteriori error estimates for the finite element method. Internat. J. Numer. Meth. Engrg., 12(1978) 1597-1615. (L2)
- I. Babuška and W. C. Rheinboldt. Error estimates for adaptive finite element computations. SIAM J.Numer. Anal., 15(1978), 736-754. (L2)
- 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. (L2)
- I. Babuška and T. Strouboulis. *The finite element method and its reliability*. The Clarendon Press, Oxford University Press, New York, 2001. (L2)
- 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. (L2)
- 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. (L2)

- 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. (L2)

- 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. (L2)

- I. Babuska and G. N. Gatica, On the mixed finite element method with Lagrange multipliers. Numer. Meth. Partial Diff. Eq. **19**(2) (2003), 192–210.

- C. Baiocchi and A. Capelo. *Variational and quasivariational inequalities. Applications to free boundary problems.* Wiley and Sons, New York, 1984. (L11)

- R. E. Bank and R. K. Smith. Mesh smoothing using a posteriori error estimates, *SIAM J. Numer. Anal.*, 34(1997), 3, 979-997. (L2, mesh adaptation minimization by means of an integral of $\int \nabla(u - u_h)$ where u is replaced by a polynomial interpolation)

- R. E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations, *Math. Comput.*, 44(1985), 283-301. (L2)
- R. E. Bank and B. D. Welfert. A posteriori error estimates for the Stokes problem, *SIAM J. Numer. Anal.*, 28(1991), 591-623. (L2)
- W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations.* Birkhäuser, Berlin, 2003. (L2)
- D. Braess and R. Verfürth, A posteriori error estimators for the Raviart-Thomas element. SIAM J. Numer. Anal. **33**(6) (1996), 2431–2444.
- J. H. Brandts, Superconvergence and a posteriori error estimation for triangular mixed finite elements. Numer. Math. **68**(3) (1994), 311–324.
- 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. (L2)

- 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. (L2)
- A. Bermúdez, R. Durán and R. Rodríguez. Finite element analysis of compressible and incompressible fluid-solid systems, *Math. Comput.*, 67(1998), 221, 111-136. (L7)
- B. Boroomand and O. C. Zienkiewicz, Recovery by equilibrium in patches (REP), *Int. J. Numer. Meth. Engrg.*, 40(1997), 137-164. (L2)
- B. Boroomand and O. C. Zienkiewicz, An improved REP recovery and the effectivity robustness test, *Int. J. Numer. Meth. Engrg.*, 40(1997), 3247-3277. (L2, variants of gradient recovery method)
- 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. (L2, a posteriori L_∞ estimates)
- D. Braess. *Finite elements.* Cambridge University Press, Cambridge, 1997. (L1,2)

- 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. (L1)
- 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. (L1)
- H. Brezis. Problémes unilatéraux, *J. Math. Pures Appl.*, 9(1971, 1, 1-168. (L11)
- Brezis H., Kinderlehrer D. The smoothness of solutions to nonlinear variational inequalities, *Indiana Univ. Math J.*, 23(1974), 831-844. (L11)
- H. Brezis and M. Sibony. Equivalence de deux inéquations variationnelles et applications, *Arch. Rat. Mech. Anal.*, 41(1971), 254-265. (L11)
- 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. (L11)

- 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. (L9)
- F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*, Springer Series in Computational Mathematics, 15, New York, 1991. (L9)
- 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. (L11)
- 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. (L11)
- I. G. Bubnov. *Selected Works.* Sudpromgiz, Leningrad (1956) (in Russian). (L1)
- 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. (L11)

- 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. (L2)
- 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. (L2)
- C. Carstensen. A posteriori estimate for the mixed finite element method, *Math. Comput.*, 66(1997), 218, 465-476. (L7, residual method for mixed approximations)
- C. Carstensen. Quasi-interpolation and a posteriori error analysis of finite element methods, *Mathematical Modelling in Numerical Analysis*, 33(1999), 6, 1187-1202. (L2, modification of Clement's interpolation operator, advanced residual-based a posteriori estimates)
- C. Carstensen, A posteriori error estimate for the mixed finite element method. Math. Comp. **66**(218) (1997), 465–476.
- C. Carstensen and G. Dolzmann, A posteriori error estimates for mixed FEM in elasticity. Numer. Math. **81** (1998), 187–209.

- 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. (L2, a posteriori estimates by post–processing)
- C. Carstensen and S. A. Funken. Fully reliable localized error control in the FEM, *SIAM J. Sci. Comput.*, 21(2000), 4, 1465-1484. (L2)
- 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. (L2, detailed investigation of the explicit residual method)
- 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. (L2)
- 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. (L2)

- Z. Chen and R. H. Nochetto. Rezidual type a posteriori error estimates for elliptic obstacle problems, *Numer. Math.*, 84(2000), 527-548. (L2,11)
- 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. (L2,7,9)
- 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. (L11)
- S. S. Chow. Finite element error estimates for non-linear elliptic equations of monotone type, *Numer. Math.*, 54(1989), 373-393. (L1, energy a priory estimates for problems with monotone operators)
- 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. (L2)
- P. G. Ciarlet. *The finite element method for elliptic problems*. North Holland, New York, Oxford, 1978. (L1)

- Ph. Clément. Approximations by finite element functions using local regularization, *RAIRO Anal. Numér.*, 9(1975), R-2, 77-84. (L2)
- *Funktionanalysis und numerische mathematik*. Springer-Verlag, Berlin, 1964. (L1,10)
- 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. (L2)
- R. Courant. Variational methods for some problems of equilibrium and vibration, *Bulletin of AMS*, 49(1943), 1-23. (L1, FEM origin)
- Computational Methods in the mechanics of Fracture, Ed. S.N. Atluri, North–Holland, New York, 1986. (L1, discussion on the reliability of computer simulation)
- P. Destuynder and B. Métivet. Explicit error bounds of a conforming finite element method, *Math. Comput.*, 68(1999), 1379-1396. (L2)

- W. Dörfler and M. Rumpf. An adaptive strategy for elliptic problems including a posteriori controlled boundary approximation, *Math. Comput.*, 67(1998), 224, 1361-1362. (L2, a posteriori estimates taking into account the influence of boundary discretization on the error)
- 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. (L2)
- R. Duran and R. Rodriguez. On the asymptotic exactness of Bank–Weiser's estimator, *Numer. Math.*, 62(1992), 297-303. (L2)
- G. Duvant and J.-L. Lions. *Les inequations en mecanique et en physique*, Dunod, Paris, 1972. (L1,11)
- I. Ekeland and R. Temam. *Convex analysis and variational problems*. North-Holland, Amsterdam, 1976. (L4)
- 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. (L10)

- K. Eriksson, D. Estep, P. Hansbo and C. Johnson. *Computational differential equations*. Cambridge University Press, Cambridge, 1996. (L1)
- K. Eriksson and C. Johnson. An adaptive finite element method for linear elliptic problems, *Math. Comput.*, 50(1988), 361-383. (L2)
- 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. (L2)
- R. E. Ewing. A posteriori error estimation, *Comput. Meth. Appl. Mech. Engrg.*, 82(1990),1-3, 59-72. (L2)
- 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. (L2)
- R. S. Falk. Error estimates for the approximation of a class of variational inequalities, *Math. Comput.*, 28(1974), 963-971. (L11)
- M. Feistauer. *Mathematical methods in fluid dynamics.* Longman, Harlow, 1993. (L1,7)

- W. Fenchel. On the conjugate convex functions, *Canad. J. Math.* 1(1949), 73-77. (L4)
- 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. (L1)
- Fraejs B. de Veubeke. A conforming finite element for plate bending, *Internat. J. Solids and Structures* 4(1968), 95-108. (L1)
- G. E. Forsythe, M. A. Malcolm and C. B. Moler. *Computer methods for mathematical computations*, Prentice-Hall, Englewood Cliffs, New York, 1977. (L1)
- A. Friedman. *Variational principles and free-boundary problems*. Wiley and Sons, New York, 1982. (L11)
- D. A. Field. Laplacian smoothing and Delanay triangulations, *Commum. Appl. Numer. Methods*, 4(1988), 709-712. (L1,2)

- 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), no. 1, 19–26. (L7)
- L. P. Franca and R. Stenberg. Error analysis of Galerkin least squares methods for the elasticity equations. *SIAM J. Numer. Anal.* 28 (1991), no. 6, 1680–1697. (L6)
- 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. Neittaanmäk and S. Repin. On computational properties of a posteriori error estimates based upon the method of duality error majorants. In *Proc. 5th European Conference on Numerical Mathematics and applications, Prague 2004*, 346–357.
- H. Gaevskii, H., K. Gröger, and K. Zacharias. *Nichtlineare Operator-gliechungen and Operatordifferentialgleiichungen.* Akademie-Verlag, Berlin, 1974.

- A. Gaevskaya and S. Repin. A posteriori error estimates for approximate solutions of linear parabolic problems. *Differential Equations*, 41 (2005), 7, 970–983. (L3)

- 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). (L1)

- L. Gallimard, P. Ladev eze and J. P. Pelle. Error estimation and adaptivity in elastoplasticity, *Int. J. Numer. Meth. Engrg.*, 39(1996), 189-217. (L2, time-dependent material models, error in Drucker inequality)

- G. Gatica and M. Maischak, A posteriori error estimates for the mixed finite element method with Lagrange multipliers. Numer. Meth. Partial Diff. Eq., to appear.

- 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.
- E. Gorshkova and S. Repin. Error control of the approximate solution to the Stokes equation using a posteriori error estimates of functional type.

In *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS 2004, P.Neittaanmäki, T. Rossi, K. Majava and O. Pironeau (eds.), O. Nevanlinna and R. Rannacher (assoc. eds.), Jyväskylä, 24-28 July, 2004* (electronic).

W. Han. Finite element analysis of a holonomic elasto-plastic problem, *Numer. Math.*, 60(1992), 493-508. (L1,2)

W. Han. A posteriori error analysis for linearization of nonlinear elliptic problems and their discretization, *Math. Meth. Appl. Sci.* , 17(1994), 487-508. (L2)

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. (L9)

P. Hansbo. Generalized Laplacian smoothing of unstructured grids, *Commun. Numer. Methods Eng.*, 11(1995), 455-464. (L2)

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. (L1)

- I. Hlaváŭek and M. Křižek. On a superconvergence finite element scheme for elliptic systems. I. Dirichlet boundary conditions. *Aplikace Matematiky*, 32(1987), No.2, 131-154. (L2)
- R. H. W. Hoppe. Mortar finite elements in R^3, *East-West J. Numer. Math.*, 7(1999), 3, 159-173. (L1,9)
- R. H. W. Hoppe and R. Kornhuber. Adaptive multilevel methods for obstacle problems. *SIAM J. Numer. Anal.*, 31(1994), 301-323. (L11)
- 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. (L2)
- A.D. Ioffe and V.M. Tikhomirov. *Theory of extremal problems. Studies in Mathematics and its Applications, 6.* North-Holland, Amsterdam-New York, 1979. (L4)
- J.L. Jensen. Sur les fonctions convexes et les inegalités entre les valeurs moyennes, *Acta Math.*, 30(1906), 175-193. (L4)

- Jin Qi-nian. Error estimates of some Newton-type methods for solving nonlinear inverse problems in Hilbert scales, *Inverse Problems*, 16(2000), 187-197. (L10)
- C. Johnson. *Numerical Solution of Partial Differential Equations by the Finite Element Methods*, Cambridge University Press, Cambridge, 1987. (L1)
- C. Johnson. On Finite Element Methods for Plasticity Problems.- *Numer. Math.*, 26(1976), 79-84. (L1)
- C. Johnson. Adaptive finite element methods for diffusion and convection problems, *Comput. Methods Appl. Mech. Engrg.*, 82(1990), 301-322. (L1,2)
- C. Johnson and P. Hansbo. Adaptive finite elements in computational mechanics, *Comput. Methods Appl. Mech. Engrg.* 101(1992), 143-181. (L2)
- C.Johnson and B.Mercier. Some equilibrium finite element methods two-dimensional elasticity problems, *Numer. Math.*, 30(1978), 101-116. (L1)

- C. Johnson and R. Rannacher. On error control in computational fluid dynamics (CFD) Preprint 1994–07, Chalmers University of Technology, Göteborg. (L2,7)
- C. Johnson, R. Rannacher and M. Boman. Numerics in hydrodynamic stability. Towards error control in CFD, *SIAM J. Numer. Anal.*, 32(1995), 1058-1079. (L2,7)
- 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. (L2)
- 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. (L2)
- H. Kavarada. *Numerical problems for free surface problems by means of penalty*. Lecture Notes in Mathematics, No 704, Springer-Verlag, 1979. (L1)
- H. Kardestuncer and D. H. Norrie ed. *Finite Element Handbook*. McGraw–Hill, New York, 1987. (L1,2)

- 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. (L2)
- D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications*. Academic Press, New York, 1980. (L11)
- 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. (L2)
- A. N. Kolmogorov and S. V. Fomin. *Introductory real analysis*. Dover Publications, Inc., New York, 1975. (L1,10)
- R. Kornhuber. A posteriori error estimates for elliptic variational inequalities, *Comput. Math. Appl.*, 31(1996), 49-60. (L2,11)
- 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. (L2)

- 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. (L1,2)
- M. Křížek and P. Neittaanmäki. *Finite Element Approximations of Variational Problems and Applications*. Wiley and Sons, New York, 1990. (L1,2)
- M. Křížek and P. Neittaanmäki. Superconvergence phenomenon in the finite element method arising from averaging of gradients *Numer. Math.*, 45(1984), 105-116. (L2)
- Yu.Kuznetsov and S.Repin. Convergence analysis and error estimates for mixed finite element method on distorted meshes. *Journal of Numerical Mathematics*, Vol.13, No.1, 2005, 22–51. (L9)
- Yu. Kuznetsov and S. Repin. New mixed finite element method on polygonal and polyhedral meshes. *Russian J. Numer. Anal. Math. Modeling*, 18(2003), no. 3, 261–278. (L9)

- Yu. Kuznetsov and S. Repin. Mixed finite element method on polygonal and polyhedral meshes. In *Proc. 5th European Conference on Numerical Mathematics and applications, Prague 2004*, 615–622. (L9)
- P. Ladev eze and D. Leguillon. Error estimate procedure in the finite element method and applications, *SIAM J. Numer. Anal.*, 20(1983), 485-509. (L2)
- P. Ladev eze, J.-P. Pelle and Ph. Rougeot. Error estimation and mesh optimization for classical finite elements, *Engineering Computations*, 8(1991), 69-80. (L2)
- 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. (L2)
- O. A. Ladyzhenskaya. *Mathematical problems in the dynamics of a viscous incompressible fluid*. Nauka, Moscow, 1970 (in Russian). (L1,7)
- O. A. Ladyzhenskaya, *The boundary value problems of mathematical physics*. Springer-Verlag, New York, 1985. (L1)

- O. A. Ladyzhenskaya, On mosified Navier–Stokes equations for large velocity gradients, *Zapiski Nauchnych Seminarov LOMI*, 7(1968), 126-154. (L7)
- O. A. Ladyzenskaja 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). (L7)
- O. A. Ladyzhenskaya and N. N. Uraltseva. *Linear and Quasilinear Elliptic equations*, Academic Press, New York, 1968. (L1)
- M. Lonsing and R. Verfürth. A posteriori error estimators for mixed finite element methods in linear elasticity. Numer. Math. **97** (2004), 757–778.
- 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). (L1,2)

- 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. (L10)
- J.-L. Lions and E. Magenes. Problḿes aux limites non homogénes et applications. Dunod, Paris, 1968 (L1).
- 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. (L2)
- 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. (L1)
- 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. (L2)

- P. Meyer. A unifying theorem on Newton's method, *Numer. Funct. Anal. Optim.* 13(1992), no. 5-6, 463-473. (L10)
- S. G. Mikhlin. *Variational methods in mathematical physics*. Pergamon, Oxford, 1964. (L1,2)
- S. G. Mikhlin. *Error Analysis in Numerical Processes* Wiley and Sons, Chiester–New York, 1991. (L2)
- S. G. Mikhlin. *Constants in some inequalities of analysis*. Wiley and Sons, Chiester–New York, 1986. (L1,3)
- P.Mosolov and V.Myasnikov. *Mechanics of rigid plastic bodies*, Nauka, Moscow, 1981 (in Russian). (L2)
- 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. (L6)
- P. Neittaanmäki and M. Křížek. On $0(h^4)$ superconvergence of piecewise bilinear FE-approximations. In Teubner-Texte Math. 107, Teubner, 1988, 250-255. (L2)

- P. Neittaanmäki and S. Repin. *Reliable methods for computer simulation. Error control and a posteriori estimates*, Elsevier, New York, 2004. (L6,7,8)
- 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. (L6)
- J. A. Nitsche and A. H. Schatz. Interior estimates for Ritz-Galerkin Methods, *Math. Comput.*, 28(1974), 128, 937-958. (L1)
- 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. (L2)
- 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. (L2)

- J. T. Oden, S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element method, *Comput. Math. Appl.*, 41, 735-756, 2001. (L2)
- J. T. Oden and J. N. Reddy. *Mathematical theory of finite elements*. Wiley and Sons, 1976. (L1)
- 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. (L2,7)
- 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). (L2, superconvergence)
- A. Ostrowski. Les estimations des erreurs a posteriori dans les procédés itératifs, *C.R. Acad.Sci. Paris Sér. A–B*, 275(1972), A275-A278. (L10)
- D.R.J.Owen and E.Hinton. *Finite Elements in Plasticity Theory: Theory and Practice*. Prinerdge Press, Swansea, 1980. (L1)

- C. Padra, A posteriori error estimators for nonconforming approximation of some quasi-newtonian flows, *SIAM J. Numer. Anal.*, 34(1997), 1600-1615. (L2,7)

- 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. (L2)

- F. Potra. Sharp error bounds for a class of Newton-like methods, *Libertas Math.* 5(1985), 71-84. (L10)

- 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. (L2)

- 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. (L2)

- W. Prager and J. L. Synge. Approximation in elasticity based on the concept of function space, *Quart. Appl. Math.* 5(1947), 241-269. (L2)
- R. Rannacher. Zur L^{∞}-Konvergenz linear finite elemente beim Dirichlet Problem, *Math. Z.*, 149(1976), 69-77. (L1)
- 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. (L1)
- R. Rannacher and R. Scott. Some optimal error estimates for piecewise linear finite element approximations, *Math. Comput.*, 38(1982), 158, 437-445. (L1)
- 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. (L2)
- 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. (L2)

- P.-A. Raviart and J.-M. Thomas. A mixed finite element for second order elliptic problems. In Mathematical Aspects of Finite Element Methods, eds. I. Galligani and E. Magenes, Springer-Verlag, Berlin, pp. 292–315, 1977.
- W. C. Rheinboldt. On a theory of mesh-refinement processes, *SIAM J. Numer. Anal.*, 17(1980), 766-778. (L2)
- 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). (L3,6,12)
- 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). (L12)
- S. Repin. A posteriori error estimation for nonlinear variational problems by duality theory. *Zapiski Nauchnych Seminarov POMI*, 243(1997) 201-214. (L3,6,12)

- 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. (L12)
- 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. (L3,6)
- S. Repin. A unified approach to a posteriori error estimation based on duality error majorants, *Mathematics and Computers in Simulation*, 50(1999), 313-329. (L6)
- S. Repin. A posteriori error estimation for variational problems with uniformly convex functionals, *Math. Comput.*, 69(230), 2000, 481-500. (L6,12)
- 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)*) (L6,12)

- 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. (L6)
- 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. (L11)
- S. Repin. Aposteriori estimates for the Stokes problem, *Journal of Math. Sciences* 109 (2002), 5, 1950-1964. (L7)
- 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. (L7)
- S. Repin. Estimates for errors in two-dimensional models of elasticity theory, *J. Math. Sci. (New York)*, 106 (2001), no. 3, 3027-3041. (L12)
- S. Repin. Local a posteriori estimates for the Stokes problem. *Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 318 (2004), Kraev. Zadachi Mat. Fiz. i Smezh. Vopr. Teor. Funkts.* 35, 233–245, 312–313. (L7)

- S. Repin. A posteriori estimates in local norms. Problems in mathematical analysis. No. 29. J. Math. Sci. (N. Y.) 124 (2004), no. 3, 5026–5035. (L12)
- S. Repin. Functional type a posteriori error estimates for approximate solutions of problems with incompressibility condition. In *European Congress on Computational Methods in Applied Sciences and Engineering, ECCOMAS 2004, P.Neittaanmäki, T. Rossi, K. Majava and O. Pironeau (eds.), O. Nevanlinna and R. Rannacher (assoc. eds.), Jyväskylä, 24-28 July, 2004* (electronic). (L7)
- S. Repin. Estimates of deviations from exact solutions for some boundary–value problems with incompressibility condition. *Algebra and Analiz*, 16(2004), 5, 124–161 (in Russian). (L7)
- S. Repin. A posteriori error estimates taking into account indeterminacy of the problem data. *Russian J. Numer. Anal. Math. Modelling*, 18 (2003), no. 6, 507–519. (L8)
- Repin, S. I.; Frolov, M. E. An estimate for deviations from the exact solution of the Reissner-Mindlin plate problem. (Russian) Zap. Nauchn.

Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI) 310 (2004), Kraev. Zadachi Mat. Fiz. i Smezh. Vopr. Teor. Funkts. 34, 145–157, 228.

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). (L6)

S. Repin and A. Smolianski. Functional-type a posteriori error estimates for mixed finite element methods. *Russian J. Numer. Anal. Math. Modelling* 20 (2005), no. 4, 365–382. (L9)

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. (L6)

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. (L6)

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), no. 4, 1435–1451. (L12)

- S. Repin, S. Sauter and A. Smolianski. A Posteriori Control of Dimension Reduction Errors on Long Domains. *Proceedings in Applied Mathematics and Mechanics*, 4, No. 1, 714–715 (2004). (L12)
- S. Repin, S. Sauter and A. Smolianski. A posteriori error estimation for the Poisson equation with mixed Dirichlet/Neumann boundary conditions. Proceedings of the 10th International Congress on Computational and Applied Mathematics (ICCAM-2002). *J. Comput. Appl. Math.* 164/165 (2004), 601–612. (L6)
- S. Repin, S. Sauter and A. Smolianski. A posteriori estimation of dimension reduction errors. In *Proc. 5th European Conference on Numerical Mathematics and applications, Prague 2004*, 717–725. (L12)
- 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. (L12)
- 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. (L12)

- W. Ritz. Über eine neue Methode zur Lözing gewisser Variationsprobleme der Mathematischen Physics, *J. Reine Angew. Math.*, 135(1909), 1-61. (L1)
- R. T. Rockafellar. *Convex analysis*, Princeton Univ. Press, 1970. (L4)
- R. Rodrigues. Some remarks on Zienkiewicz–Zhu estimator, *Numer. Methods for PDE*, 10(1994), 625-635. (L2)
- W. Rudin. *Functional analysis*, McGraw-Hill, 1973. (L1)
- J.L.Segerlind. *Applied finite element analysis*, Wiley and Sons , 1976. (L1)
- 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. (L1)
- A. H. Schatz and L. B. Wahlbin. Interior maximum norm estimates for finite element methods, *Math. Comput.*, 31(1977), 414-442. (L1)

- 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. (L1)
- 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. (L1)
- C. Schwab. A-posteriori modelling error estimation for hierarchic plate models, *Numer. Math.*, 74(1996), 221-259. (L12)
- 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. (L1)
- 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. (L2)
- 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. (L2)

- M. Schultz and W. Wendland. *On an adaptive finite element method for elasto-plastic deformation with hardening.* Preprint 1997/39, SFB 259, Stuttgardt, 1997. (L2)

- 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). (L1)

- E. Stein and S. Ohnimus. Coupled model- and solution-adaptivity in the finite element method, *Comput. Methods Appl. Mech. Engrg.*, 150(1997), 327-350. (L2, local residual estimators for 1D and 2D models, equilibrium method).

- 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. (L2)

- R. Stenberg. Some new families of finite elements for the Stokes equations. *Numer. Math. 56 (1990), no. 8, 827–838.* (L7, Taylor-Hood elements, inf-sup condition)
- R. Stenberg. Error analysis of some finite element methods for the Stokes problem. *Math. Comp.* 54 (1990), no. 190, 495–508. (L7)
- G. Strang and G. Fix. *An analysis of the finite element method.* Prentice Hall, Englewood Cliffs, 1973. (L1)
- 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. (L2)
- 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. (L2)
- R. Temam. Problèmes mathématique en plasticité. Bordas, Paris, 1983. (L4)

- K. Tsuruta and K. Ohmori. A posteriori error estimation for Volterra integro-differential equations, *Mem. Numer. Math. No. 3* 1976, 33-47. (L10)

- N.N. Uraltseva. On the regularity of solutions to variational inequalities *Uspekhi Mat. Nauk*, 42(1987), 6, 151-174 (in Russian). (L11)

- R.S. Varga. *Matrix Iterative Analysis.* Prentice–Hall, Englewood Cliffs, New Jersey, 1962. (L10)

- A. Veeser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems, *SIAM J. Numer. Anal.*, 39(2001), 146-167. (L2,11)

- R. Verfürth. A posteriori error estimators for the Stokes equations, *Numer. Math.*, 55(1989), 309-326. (L2,7)

- R. Verfürth. A posteriori error estimates for nonlinear problems. Finite element discretisations of elliptic equations, *Math. Comput.* 62(1994) 445-475. (L2)

- R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques* Wiley and Sons, Teubner, New-York, 1996.

(L2, mathematical style presentation of the residual and averaging methods)

- 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. (L2, general "residual" scheme for parabolic problems)
- A. Vesser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems, *SIAM J. Numer. Anal.*, 39(2001), 1, 146-167. (L2,11)
- M. I. Vishik . The method of orthogonal projections for selfadjoint equations , *Sov. Math. Dokl* , 1947, 56. (L2)
- M. Vohralik. Equivalence between mixed finite element and multi-point finite volume methods. C.R. Acad. Sci. Paris, Ser. I, **339** (2004), 525–528.
- B.I. Wohlmuth and R. H. W. Hoppe. A comparison of a posteriori error estimators for mixed finite element discretizations by Raviart-Thomas elements. Math. Comp. **68**(228) (1999), 1347–1378.

- A. Younes, P. Ackerer and G. Chavent. From mixed finite elements to finite volumes for elliptic PDEs in two and three dimensions. Int. J. Numer. Meth. Engng. **59** (2004), 365–388.
- L. B. Wahlbin. *Superconvergence in Galerkin Finite Element Methods*, Lecture Notes in Mathematics, No 1605, Springer-Verlag, 1995. (L2)
- 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. (L2)
- H. Weil. The method of orthogonal projections in potential theory, *Duke Math. J.*, 7(1940), 411-444. (L2)
- 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. (L2)
- 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. (L2)

- On classes of summable functions and their Foutier series, *Proc. Roy. Soc. Ser. A*, 87(1912), 225-229. (L4)
- 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. (L2)
- E. Zeidler. *Nonlinear functional analysis and its applications. I. Fixed-point theorems.* Springer-Verlag, New York, 1986. (L10)
- O. C. Zienkiewicz. Achievements and some unsolved problems of the finite element method, *Int. J. Numer. Meth. Engrg.*, 47(2000), 9-28. (L1,2)
- 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. (L2)
- O. C. Zienkiewicz and K. Morgan. *Finite elements and approximation*. Wiley and Sons, New York, 1983. (L1)

- 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. (L1)
- 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. (L2)
- O. C. Zienkiewicz and J. Z. Zhu. Adaptive techniques in the finite element method, *Commun. Appl. Numer. Methods*, 4(1988), 197-204. (L2)
- 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. (L2)
- J. Z. Zhu. A posteriori error estimation - the relationship between different procedures, *Comput. Methods Appl. Mech. Engrg.*, 150(1997), 411-422. (L2)
- 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. (L2)