

A POSTERIORI ERROR ESTIMATION OF MODELING ERRORS.

S. Repin

Saint Petersburg Department of V.A. Steklov Institute of Mathematics

Modeling errors

We have two different meanings of the term **modeling error**

- I. Original Problem \mathcal{P} has solution u (or the pair (u, p^*)),
Modified Problem $\tilde{\mathcal{P}}$ has solution \tilde{u} (or the pair (\tilde{u}, \tilde{p}^*))

$$\|u - \tilde{u}\| = \text{error generated by the difference of models}$$

- II. \mathcal{P} is a mathematical model of some process, object, or phenomenon.
Values of the function u are compared with **experimental data**. The difference is evaluated and shows the applicability of \mathcal{P} .

Mathematical models with small parameter

$$\text{Problem } \mathcal{P} \leftrightarrow \text{Problem } \mathcal{P}_\epsilon$$

- Penalty method:

$$J(v) + \frac{1}{\epsilon} \Psi(v)$$

- Regularisation method:

$$\epsilon \|\Lambda u\|^2 + \dots$$

- Dimension reduction:

$$\Omega \subset (0, d_1) \times (0, d_2) \times (0, \epsilon), \quad d_1, d_2 \gg \epsilon$$

- Homogenization:

$$\epsilon - \boxplus \text{ cell parameter}$$

In fact, approximation errors can be also viewed as modeling errors.

Problem \mathcal{P} is defined for functions in V .

Problem \mathcal{P}_h is defined for functions in $V_h \subset V$.

h is a small parameter.

$u - u_h$ is a modeling error
arising due to replacing a functional (continual) model
by its finite dimensional analog.

Original Problem \mathcal{P} has solution u (or the pair (u, p^*)),
Modified Problem \mathcal{P}_ϵ has solution u_ϵ (or the pair $(u_\epsilon, p_\epsilon^*)$)

$$\|u - u_\epsilon\| = (\leq \geq) ?$$

or/and

$$\|p^* - p_\epsilon^*\| = (\leq \geq) ?$$

Classical approach: a priori error analysis

$$\epsilon \rightarrow 0$$

- prove that $u_\epsilon \rightarrow u$ in some suitable space V .
- establish the rate of convergence $\|u - u_\epsilon\|_V \leq C\epsilon^m$

A posteriori approach

$$\epsilon \text{ is fixed} \quad \|u - u_\epsilon\|_V \leq M(u_\epsilon, \mathcal{D})$$

\mathcal{D} all known problem data: domain, coefficients, boundary conditions and other parameters which we have at hand;

M is directly computable;

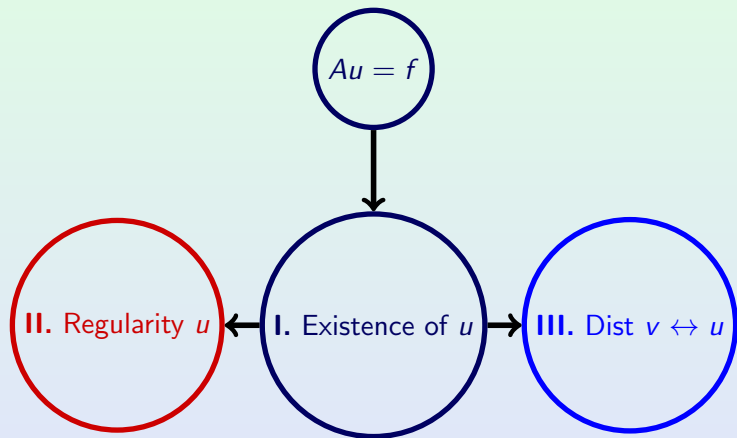
$M \rightarrow 0$ if $\epsilon \rightarrow 0$;

M is consistent, i.e., it has the same convergence rate as the a priori estimate.

Getting M with such properties may be a challenging problem!

Solving it requires certain elaboration of methods that are used for analysis of PDEs.

3 principal problems in analysis of PDEs



QUALITATIVE ANALYSIS

QUANTITATIVE ANALYSIS

Problems I, II, and III arise for ANY problem associated with a differential or an integral equation or a system of equations. Problem III arises because u is unavailable and in quantitative analysis we replace it by some v

To solve Problem III we need to have *estimates of deviations from u*

$$M_{\ominus}(v, \mathcal{D}) \leq \mu(v - u) \leq M_{\oplus}(v, \mathcal{D})$$

\mathcal{D} – problem data, μ – error measure

M_{\ominus} – error minorant, M_{\oplus} – error majorant

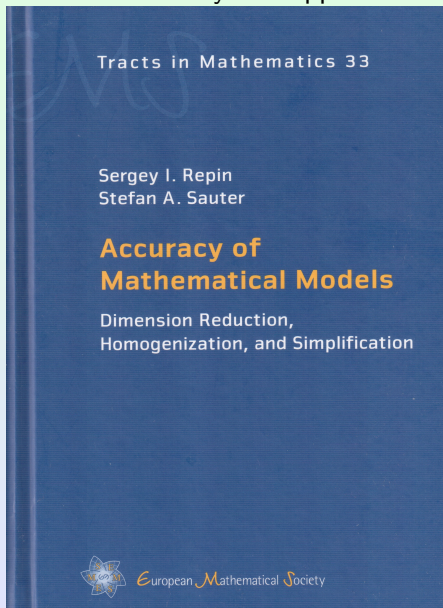
These estimates are valid for ANY admissible function. We can use them to derive bounds of **modeling errors**.

We consider the exact solution of a simplified model $\hat{\mathbf{u}}$ as an approximation of u and set $\mathbf{v} = \hat{\mathbf{u}}$ into the deviation estimate. Then $M_{\oplus}(\hat{\mathbf{u}}, \mathcal{D})$ and $M_{\ominus}(\hat{\mathbf{u}}, \mathcal{D})$ yield a bounds of the modeling error.

.....
This method was applied to analysis of errors arising due to

- Simplification
- Penalization
- Linearization
- Dimension reduction
- Homogenization
- Also, the method is applicable to analysis of surrogate models generated by [Deep Neural Networks](#)

A systematic exposition of the theory and applications:



The basic theory

In principle, the ultimate goal of the error control is to deduce for a problem the **error identity**

$$\mu(\textit{error}) = \mathbb{F}(\mathcal{D})$$

μ – error measure,
 \mathbb{F} depends on the problem data and known functions

Sometimes it is easy to do, sometimes not...

Simple example

$$\begin{aligned} -\operatorname{div} \mathbf{A} \nabla u + \delta u &= f && \text{in } \Omega \\ u &= u_0 && \text{on } \Gamma. \end{aligned}$$

where $f \in L^2(\Omega)$ is a given function and $\mathbf{A}(x)$ is a positive definite matrix. Let $v \in V_0 + u_0$, $V_0 := \mathring{H}^1(\Omega)$ and $\mathbf{y}^* \in H(\Omega, \operatorname{div})$ be approximations. In this case, we can write error identity for $\mathbf{e} = v - u$ and $\mathbf{e}^* = \mathbf{y}^* - \mathbf{p}^*$

$$\mu(\mathbf{e}) + \mu^*(\mathbf{e}^*) = \|\mathbf{A} \nabla v - \mathbf{y}^*\|_{\mathbf{A}^{-1}}^2 + \frac{1}{\delta} \int_{\Omega} |\operatorname{div} \mathbf{y}^* - \delta v + f|^2 dx.$$

where error measures are certain (combined) norms of \mathbf{e} and \mathbf{e}^* :

$$\mu(\mathbf{e}) + \mu^*(\mathbf{e}^*) = \|\nabla \mathbf{e}\|_{\mathbf{A}}^2 + \delta \|\mathbf{e}\|^2 + \|\mathbf{e}^*\|_{\mathbf{A}^{-1}}^2 + \frac{1}{\delta} \|\operatorname{div} \mathbf{e}^*\|^2$$

Comment: Possible application to singularly perturbed problems:

$$\begin{aligned} -\epsilon \operatorname{div} A \nabla u_\epsilon + u_\epsilon &= g && \text{in } \Omega \\ u &= u_0 && \text{on } \Gamma. \end{aligned}$$

Here $\delta = \frac{1}{\epsilon}$, $f = \frac{1}{\epsilon}g$. Let v_ϵ be an approximation of u_ϵ and \mathbf{y}_ϵ^* be approximation of $\mathbf{p}_\epsilon^* = A \nabla u_\epsilon$. Then

$$\begin{aligned} \|\nabla(v_\epsilon - u_\epsilon)\|_A^2 + \frac{1}{\epsilon} \|v_\epsilon - u_\epsilon\|^2 + \|\mathbf{y}_\epsilon^* - \mathbf{p}_\epsilon^*\|_{A^{-1}}^2 + \epsilon \|\operatorname{div}(\mathbf{y}_\epsilon^* - \mathbf{p}_\epsilon^*)\|^2 \\ = \|A \nabla v_\epsilon - \mathbf{y}_\epsilon^*\|_{A^{-1}}^2 + \epsilon \int_{\Omega} \left| \operatorname{div} \mathbf{y}_\epsilon^* - \frac{1}{\epsilon} v_\epsilon + \frac{1}{\epsilon} g \right|^2 dx. \end{aligned}$$

Multiply by ϵ

$$\begin{aligned} \epsilon \|\nabla(v_\epsilon - u_\epsilon)\|_A^2 + \|v_\epsilon - u_\epsilon\|^2 + \epsilon \|\mathbf{y}_\epsilon^* - \mathbf{p}_\epsilon^*\|_{A^{-1}}^2 + \epsilon^2 \|\operatorname{div}(\mathbf{y}_\epsilon^* - \mathbf{p}_\epsilon^*)\|^2 \\ = \epsilon \|A \nabla v_\epsilon - \mathbf{y}_\epsilon^*\|_{A^{-1}}^2 + \int_{\Omega} |\epsilon \operatorname{div} \mathbf{y}_\epsilon^* - v_\epsilon + g|^2 dx \end{aligned}$$

v_ϵ and \mathbf{y}_ϵ^* can be constructed by any method.

Estimates of deviations are derived by two methods:

Calculus of variation

For elliptic equations, variational problems.

Transformations of integral relations that define generalised solutions of
BVPs

For non-variational problems, evolutionary problems.

.....
Blackboard comments

We discuss one class of PDEs that arises as Euler's equations associated with variational formulations

$$\inf_{v \in V} J(v), \quad J(v) = G(\Lambda w) + F(w)$$

Notation:

V, Y – reflexive Banach spaces,

$G : Y \rightarrow \mathbb{R}_+$: convex, continuous, coercive functional vanishing at zero element of Y ,

$F : V \rightarrow \mathbb{R}$ – convex, l.s.c. functional,

$\Lambda : V \rightarrow Y$ bounded linear operator

Pairing of spaces Y and $Y^* \Rightarrow \langle y^*, y \rangle$, V and $V^* \Rightarrow \langle v^*, v \rangle$.

$\Lambda : V \rightarrow Y$ is the differential operator (e.g., ∇ or ∇_{sym}),

$\Lambda^* : Y^* \rightarrow V^*$ is the conjugate operator (e.g., $-\text{div}$ or $-\text{Div}$):

$$\langle \Lambda^* y^*, v \rangle = \langle y^*, \Lambda v \rangle$$

e.g., for $v \in V_0$

$$\int_{\Omega} \nabla v \cdot y^* dx = - \int_{\Omega} v \text{div} y^* dx$$

This class contains:

α -Laplacian, NonNewtonian fluids, nonlinear diffusion and reaction–diffusion, Linear and physically nonlinear elasticity, Elasto–plasticity, Models with unilateral and obstacle conditions, field theory models...

Examples:

$$J(v) = \int_{\Omega} \left(\frac{1}{2} A \nabla v \cdot \nabla v - f v \right) dx \quad \text{linear diffusion}$$

$$J(v) = \int_{\Omega} \left(\frac{1}{2} L \varepsilon(v) : \varepsilon(v) - f \cdot v \right) dx - \int_{\Gamma} F \cdot v ds \quad \text{elasticity}$$

$$J(v) = \int_{\Omega} \left(\frac{1}{\alpha} |\nabla v|^{\alpha} + \delta |v|^{\beta} \right) dx - \int_{\Omega} f v dx \quad \text{nonlinear reaction–diffusion}$$

$$J(v) = \int_{\Omega} \left(\frac{\nu}{2} |\varepsilon(v)|^2 + k_* |\varepsilon(v)| - \lambda v \right) dx \quad \text{Bingham problem}$$

One more example: α - Laplacian

Let $\frac{1}{\alpha} + \frac{1}{\alpha^*} = 1, \alpha > 1, V = \overset{\circ}{W}^{1,\alpha}(\Omega), Y = L^\alpha(\Omega, \mathbb{R}^d),$
 $Y^* = L^{\alpha^*}(\Omega, \mathbb{R}^d),$

$$\Lambda = \nabla, \quad \Lambda^* = -\operatorname{div}, \quad G(y) = \frac{1}{\alpha} \int_{\Omega} |y|^\alpha dx, \quad F(v) = \int_{\Omega} fv dx$$

We arrive at the functional

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

whose minimizer satisfies the equation:

$$\operatorname{div} |\nabla u|^{\alpha-2} \nabla u + f = 0, \text{ in } \Omega, \quad u = 0 \text{ on } \Gamma$$

We need some specific notions and results in nonlinear analysis,^a

^aT. Rockafellar, J. Moreau, I. Ekeland and R. Temam...

Fenchel conjugate functional to the functional $g : X \rightarrow X^*$:

$$g^*(\zeta^*) := \sup_{\zeta \in X} \{ \langle \zeta^*, \zeta \rangle - g(\zeta) \}$$

Example: if $g(\zeta) = \frac{1}{\alpha} |\zeta|^\alpha$, then $g^*(\zeta^*) = \frac{1}{\alpha^*} |\zeta^*|^{\alpha^*}$

Compound functional is defined on $X \times X^*$. It possesses two important properties that make it a natural **error measure**.

"Sign property"

$$D_g(\zeta, \zeta^*) := g(\zeta) + g^*(\zeta^*) - \langle \zeta^*, \zeta \rangle \geq 0 !$$

and "Vanishing conditions":

$$D_g(\zeta, \zeta^*) = 0 \Leftrightarrow \zeta^* \subset \partial g(\zeta) \text{ and } \zeta \subset \partial g^*(\zeta^*)$$

Special case: quadratic energy \Rightarrow linear problems \Rightarrow standard error norms

If X is a Hilbert space so that $X = X^*$ and $g(\xi) = \frac{1}{2}\|\xi\|^2$

then $g^*(\xi^*) = \frac{1}{2}\|\xi^*\|^2$. In this case,

$$D_g(\xi, \xi^*) = \frac{1}{2}\|\xi\|^2 + \frac{1}{2}\|\xi^*\|^2 - (\xi, \xi^*) = \frac{1}{2}\|\xi - \xi^*\|^2$$

and D_g is reduced to the norm of X .

For this reason basic error identities
for linear problems (**and only for them!**) are presented
in terms of norms.

Functional J has a dual counterpart

$$\mathcal{I}^*(y^*) := -\mathcal{F}^*(-\Lambda^* y^*) - \mathcal{G}^*(y^*),$$

which generates the (dual) variational Problem \mathcal{P}^* : find $p^* \in Y^*$ such that

$$\mathcal{I}^*(p^*) = \sup_{y^* \in Y^*} \mathcal{I}^*(y^*).$$

These two problems are joined by the relation

$$\mathcal{I}^*(y^*) \leq \mathcal{J}(v) \quad \forall v \in V, y^* \in Y^*$$

Moreover, if the solutions exists, then

$$\sup_{y^* \in Y^*} \mathcal{I}^*(y^*) = \mathcal{I}^*(p^*) = \mathcal{J}(u) = \inf_{v \in V} \mathcal{J}(v)$$

The Main Error Identity for variational problems

For any pair of approximations $v \in V$ and $y^* \in Y^*$

$$\mu(v) + \mu^*(y^*) = D_G(\Lambda v, y^*) + D_F(v, -\Lambda^* y^*)$$

error measure μ

=

computable quantity

Here the error measure consists of four terms and two parts

$$\begin{aligned}\mu(v) &= D_F(v, -\Lambda^* p^*) + D_G(\Lambda v, p^*), \\ \mu^*(y^*) &= D_F(u, -\Lambda^* y^*) + D_G(\Lambda u, y^*).\end{aligned}$$

COMMENT: $D_G(\Lambda v, p^*)$ is a nonlinear measure of the distance from v to u , e.g., if G is differentiable then $D_G(\Lambda v, p^*) = D_G(\Lambda v, G'(\Lambda u))$.

It has been proven that the measure $\mu(v, y^*) := \mu(v) + \mu^*(y^*)$ is equal to the duality gap $J(v) - I^*(y^*)$, hence

$$\mu(v, y^*) = 0 \text{ iff } \{v, y^*\} \text{ is equal to } \{u, p^*\}$$

Conclusion: $\mu(v, y^*)$ is the natural error measure, which is generated by the variational (energy) formulation. It generates the strongest local topology available for approximations constructed by this method.

The identity is the key point for a posteriori analysis of numerical

approximations constructed by primal methods

$$J(u_k) \rightarrow \inf J$$

dual methods

$$I^*(p_k^*) \rightarrow \sup I^*$$

primal–dual (mixed) methods

$$L(u_k, p_k^*) \rightarrow \text{saddle point } L(u, p^*)$$

Quite similarly they are used for a posteriori analysis of modeling errors.

A very simple example, where $\mu(v, y^*)$ is easy to visualise

$$J(v) = \frac{1}{\alpha} |\kappa v|^\alpha + \frac{1}{\beta} |v|^\beta$$

$$V = Y = \mathbb{R}, \alpha, \beta > 1,$$

$u = 0$ is the minimizer.

$$G(y) = \frac{1}{\alpha} |y|^\alpha, \quad F(v) = \frac{1}{\beta} |v|^\beta, \quad \Lambda v = \kappa v, \quad \Lambda^* y^* = \kappa y^*,$$

$$G^*(y^*) = \frac{1}{\alpha^*} |y^*|^{\alpha^*}, \quad F^*(v^*) = \frac{1}{\beta^*} |v^*|^{\beta^*},$$

$$I^*(y^*) = -\frac{1}{\alpha^*} |y^*|^{\alpha^*} - \frac{|\kappa|^{\beta^*}}{\beta^*} |y^*|^{\beta^*}, \quad p^* = 0 \text{ maximizer.}$$

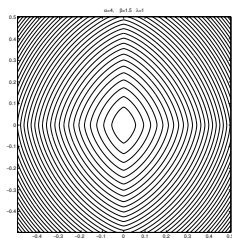
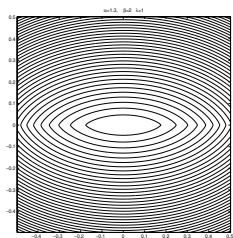
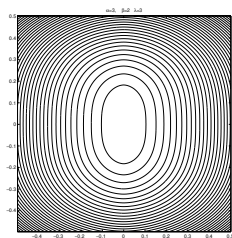
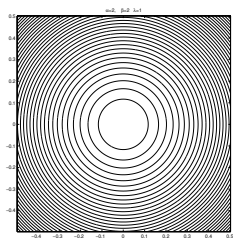
Full measure $\mu(v, y^*)$ between (v, y^*) and $(0, 0)$ induced by the problem

$$\mu(v) + \mu^*(y^*) = \frac{|\kappa|^\alpha}{\alpha} |v|^\alpha + \frac{1}{\beta} |v|^\beta + \frac{1}{\alpha^*} |y^*|^{\alpha^*} + \frac{|\kappa|^{\beta^*}}{\beta^*} |y^*|^{\beta^*}$$

is

Level lines of the measure μ for different nonlinearities

$\alpha = 2, \beta = 2, \kappa = 1$ (top left) (linear problem), $\alpha = 3, \beta = 2, \kappa = 3$ (top right),
 $\alpha = 1.3, \beta = 2, \kappa = 1$ (bottom left) and $\alpha = 4, \beta = 1.5, \kappa = 1$ (bottom right)



Important class of problems: linear functional $F = \langle \ell, v \rangle$

$$\Delta u + f = 0 \Rightarrow F(v) = \int_{\Omega} f v dx$$

Basic example:

$$F^*(v^*) = \sup_{v \in V} \langle v^* - \ell, v \rangle = \begin{cases} 0 & \text{if } v^* = \ell, \\ +\infty & \text{if } v^* \neq \ell \end{cases}$$

and, therefore,

$$F^*(-\Lambda^* y^*) = \begin{cases} 0 & \text{if } \Lambda^* y^* + \ell = 0, \\ +\infty & \text{if } \Lambda^* y^* + \ell \neq 0. \end{cases},$$

$$D_F(v, -\Lambda^* y^*) = \langle \ell, v \rangle + F^*(-\Lambda^* y^*) + \langle \Lambda^* y^*, v \rangle = \begin{cases} 0 & \text{if } \Lambda^* y^* + \ell = 0, \\ +\infty & \text{if } \Lambda^* y^* + \ell \neq 0. \end{cases}$$

For this class dual variable is restricted to:

$$q^* \in Q_{\ell}^* := \{q^* \in Y^* \mid \Lambda^* q^* + \ell = 0\}.$$

If $q^* \in Q_\ell^*$, then all compounds D_F generated by F vanish: and the identity is meaningful only if $q^* \in Q_\ell^*$.

Now, the a posteriori error identity reads

$$\mu_G(\nu) + \mu_G^*(q^*) = D_G(\Lambda \nu, q^*).$$

This leads the most general form of the so-called **hypercircle** estimate

$$\mu_G(\nu) \leq D_G(\Lambda \nu, q^*) \quad \forall q^* \in Q_\ell^*$$

Example: α - Laplacian. $V_0 = \mathring{W}^{1,\alpha}(\Omega)$, $\mathcal{J}(v) = \frac{1}{\alpha} \int_{\Omega} |\nabla u|^\alpha - \int_{\Omega} f v$

$$Q_\ell^* = \{\mathbf{q}^* \in Y^* = W^{1,\alpha^*}(\Omega) \mid \int_{\Omega} (\mathbf{q}^* \cdot \nabla w - f w) d\mathbf{x} = 0 \quad \forall w \in V_0\}$$

$$\mu_{\mathcal{G}}(v) + \mu_G^*(\mathbf{q}^*) = \int_{\Omega} \left(\frac{1}{\alpha} |\nabla v|^\alpha + \frac{1}{\alpha^*} |\mathbf{q}^*|^{\alpha^*} - \nabla v \cdot \mathbf{q}^* \right) d\mathbf{x}$$

$$\mu_{\mathcal{G}}(v) = \int_{\Omega} \left(\frac{1}{\alpha} |\nabla v|^\alpha + \frac{1}{\alpha^*} |\nabla u|^\alpha - \nabla v \cdot \nabla u |\nabla u|^{\alpha-2} \right) d\mathbf{x}$$

$$\mu_G^*(\mathbf{q}^*) = \int_{\Omega} \left(\frac{1}{\alpha} |\mathbf{p}^*|^{\alpha^*} + \frac{1}{\alpha^*} |\mathbf{q}^*|^{\alpha^*} - |\mathbf{p}^*|^{\frac{2-\alpha}{\alpha-1}} \mathbf{p}^* \cdot \mathbf{q}^* \right) d\mathbf{x},$$

Notice that $\mu_{\mathcal{G}}(v)$ differs from $\frac{1}{\alpha} \|\nabla(u-v)\|_{\alpha,\Omega}^\alpha$!

If $\alpha = \alpha^* = 2$, we arrive at the well known hypercircle identity

$$\|\nabla(v-u)\|^2 + \|\mathbf{q}^* - \mathbf{p}^*\|^2 = \|\nabla v - \mathbf{q}^*\|^2$$

Extended form of the error identity

For any $v \in V$ and $y^* \in Y^*$

$$\mu_{\mathcal{G}}(v) + \mu_{\mathcal{G}}^*(y^*) = \mathcal{D}_{\mathcal{G}}(\Lambda v, y^*) + \langle \Lambda^* y^* + \ell, v - u \rangle$$

The term $\langle \Lambda^* y^* + \ell, v - u \rangle$ can be estimated by different methods what yields various forms of M_{\oplus} and M_{\ominus} .

Example. Linear problems

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

$\mathcal{A} : Y \rightarrow Y$ is positive definite, $u, v \in V$,

\mathcal{V} is a space which norm is presented by an integral (i.e., it is computable), C is a constant in the functional inequality

$$\|w\|_{\mathcal{V}} \leq C \|\Lambda w\|_Y \quad \forall w \in \mathcal{V}.$$

For any $\beta > 1$, we have $M_{\oplus}(v, y, \beta)$

$$\|\mathbf{e}^*\|_{\mathcal{A}^{-1}}^2 + \frac{\beta - 1}{\beta} \|\Lambda \mathbf{e}\|_{\mathcal{A}}^2 \leq \|\mathcal{A} \Lambda \mathbf{v} - \mathbf{y}^*\|_{\mathcal{A}^{-1}}^2 + \beta \mathbf{C}^2 \|\ell + \Lambda^* \mathbf{y}^*\|_{\mathbf{V}}^2$$

and $M_{\ominus}(v, y, \beta)$

$$\|\mathbf{e}^*\|_{\mathcal{A}^{-1}}^2 + \frac{\beta + 1}{\beta} \|\Lambda \mathbf{e}\|_{\mathcal{A}}^2 \geq \|\mathcal{A} \Lambda \mathbf{v} - \mathbf{y}^*\|_{\mathcal{A}^{-1}}^2 - \beta \mathbf{C}^2 \|\ell + \Lambda^* \mathbf{y}^*\|_{\mathbf{V}}^2$$

There exist various modifications of these estimates.

The estimates are well verified error control tools for numerical errors, see

- S. R., S. Sauter, A. Smolianski, *Two-sided a posteriori error estimates for mixed formulations of elliptic problems*, SIAM J. Num. Analysis, (2007). [mixed FEM](#)
- R. Lazarov, S. R., S. Tomar, *Functional a posteriori error estimates for discontinuous Galerkin approximations of elliptic problems*, Numer. Methods Partial Differential Equations 25, 4, 952–971, (2009). [DG method](#)
- S. Cochez-Dhondt, S. Nicaise, S. R., *A posteriori error estimates for finite volume approximations*, Math. Model. Nat. Phenom. 4, 1, 106–122, (2009). [FV method](#)
- O. Mali, P. Nettaanmäki, S. R., *Accuracy verification methods. Theory and Algorithms*, Springer, Berlin, (2014). [FEM methods](#)
- S. Kurz, D. Pauly, D. Praetorius, S.R., · D. Sebastian, *Functional a posteriori error estimates for boundary element methods*. Numer. Math., (2021). [BEM](#)
- S. K. Kleiss and S. K. Tomar. *Guaranteed and sharp a posteriori error estimates in isogeometric analysis*. Comput. Math. Appl., 70(3),167–190, (2015). [IgA](#)

We discuss applications of the theory to Modeling Errors

Simplification of models

Typical cases:

- highly oscillating coefficients,
- complicated source terms and boundary conditions,
- domains with irregular boundaries

Given an a priori desired accuracy ϵ can we use a simpler model instead of the original one?

Simplest example: $\operatorname{div} a \nabla u + f = 0$ and $\operatorname{div} \tilde{a} \nabla \tilde{u} + f = 0$. In the variational form:

$$\mathcal{J}(u) = \inf_{v \in V_0 = H^1(\Omega)} \mathcal{J}(v), \quad \mathcal{J}(v) = \frac{1}{2} \int_{\Omega} a(x) |\nabla v|^2 d\mathbf{x} + \int_{\Omega} f v d\mathbf{x},$$

$$\tilde{\mathcal{J}}(\tilde{u}) = \inf_{v \in V_0} \tilde{\mathcal{J}}(v), \quad \tilde{\mathcal{J}}(v) = \frac{1}{2} \int_{\Omega} \tilde{a}(x) |\nabla v|^2 d\mathbf{x} + \int_{\Omega} f v d\mathbf{x},$$

where $a, \tilde{a} \in L^\infty(\Omega)$, $\tilde{a} > 0$ is much simpler (e.g., a piecewise constant function)
In this case, Λ is the gradient operator,

$$G(y) = \frac{1}{2} \int_{\Omega} a |y|^2 d\mathbf{x}, \quad G^*(y^*) = \int_{\Omega} \frac{1}{2a} |y^*|^2 d\mathbf{x}, \quad \mathbf{p}^* = a \nabla u,$$

$$\tilde{G}(y) = \frac{1}{2} \int_{\Omega} \tilde{a} |y|^2 d\mathbf{x}, \quad \tilde{G}^*(y^*) = \int_{\Omega} \frac{1}{2\tilde{a}} |y^*|^2 d\mathbf{x}, \quad \tilde{\mathbf{p}}^* = \tilde{a} \nabla \tilde{u}.$$

Hence the error identity reads

$$\begin{aligned}\mu(\tilde{u}) + \mu^*(\tilde{\mathbf{p}}^*) &= G(\nabla \tilde{u}) + G^*(\tilde{\mathbf{p}}^*) - \int_{\Omega} \nabla \tilde{u} \cdot \tilde{\mathbf{p}}^* d\mathbf{x} \\ &= \frac{1}{2} \int_{\Omega} \left((a - \tilde{a}) |\nabla \tilde{u}|^2 + \left(\frac{1}{a} - \frac{1}{\tilde{a}} \right) |\tilde{\mathbf{p}}^*|^2 \right) d\mathbf{x}.\end{aligned}$$

It implies the error identity for error of simplification

$$\mathbf{e}_{\text{mod}}^2 := \|\nabla(\mathbf{u} - \tilde{\mathbf{u}})\|_a^2 + \|\mathbf{p}^* - \tilde{\mathbf{p}}^*\|_{a^{-1}}^2 = \|\nabla \tilde{\mathbf{u}}\|_{\delta}^2,$$

where $\|\nabla \tilde{\mathbf{u}}\|_{\delta}^2 := \int_{\Omega} \delta(x) |\nabla \tilde{\mathbf{u}}|^2 d\mathbf{x}$ and $\delta(x) = \frac{(a - \tilde{a})^2}{a}$.

But the function $\tilde{\mathbf{u}}$ is unknown. How to bypass this difficulty?

We can use a straightforward way and bound the right hand side with the help of the energy estimate

$$\|\nabla \tilde{u}\|_{\Omega} \leq \frac{C_F(\Omega)}{\tilde{a}_{\ominus}} \|f\|_{\Omega}, \quad \text{where } \tilde{a}_{\ominus} := \operatorname{ess\,inf}_{x \in \Omega} \tilde{a}(x).$$

Let $\varkappa = \operatorname{ess\,sup}_{x \in \Omega} |\delta(x)|$. Then, we obtain

$$e_{\text{mod}} \leq \frac{C_F(\Omega)}{\tilde{a}_{\ominus}} \varkappa \|f\|_{\Omega}$$

This upper bound is based on L_{∞} estimates of $a - \tilde{a}$ and, in general, is too pessimistic.

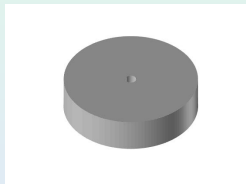
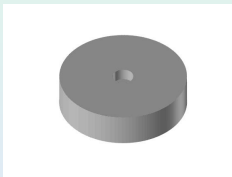
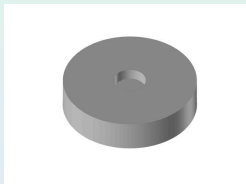
Example [removing isolated "pores"].

For $r > 0$, let B_r denote the open ball in \mathbb{R}^2 with radius r .

Let $\Omega := B_1$ and, for $0 \leq \varepsilon < 1$, let $\omega_{\varepsilon,2} := B_\varepsilon$ and $\omega_{\varepsilon,1} := \Omega \setminus \overline{\omega_{\varepsilon,2}}$.

We define positive, constant coefficients

$$a_\varepsilon(\mathbf{x}) := \begin{cases} a_1 & \text{if } x \in \omega_{\varepsilon,1}, \\ a_2 & \text{if } x \in \omega_{\varepsilon,2}. \end{cases}$$



$$\kappa = \frac{(a_1 - a_2)^2}{a_1}$$

In this case it does not depend on the pore size ε and does not tend to zero as $\varepsilon \rightarrow 0$.

But $\|\nabla(u - u_\varepsilon)\|$ may converge to zero!

Problem with boundary conditions in polar coordinates $u = g(\alpha) := \cos \alpha$.

$$u_\varepsilon(r, \alpha) = \begin{cases} \frac{2a_1 r}{a_1 + a_2} \frac{\cos \alpha}{1 - \varepsilon^2 \delta} & 0 \leq r < \varepsilon \\ (r - \varepsilon^2 \delta \frac{1}{r}) \frac{\cos \alpha}{1 - \varepsilon^2 \delta} & \varepsilon \leq r < 1 \end{cases}$$

where $\delta := \frac{a_2 - a_1}{a_2 + a_1}$. We have the estimate

$$\|\nabla u_\varepsilon - \nabla \tilde{u}\|_{p, \Omega} \leq C_{\varepsilon, \delta} \left(\frac{p}{p-1} \right)^{1/p} \frac{\|\tilde{a} - a_\varepsilon\|_{p, \Omega}}{\tilde{a} + a_1}, \quad C_{\varepsilon, \delta} := \left| \frac{1}{1 - \delta \varepsilon^2} \right|$$

For any $2 \leq p < \infty$ $\|\tilde{a} - a_\varepsilon\|_{p, \Omega}$ tends to zero and hence $\|\nabla u_\varepsilon - \nabla \tilde{u}\|_{p, \Omega}$ converges to zero as $\varepsilon \rightarrow 0$.

Remark: The situation changes significantly if the number of inclusions (pores) tend to infinity and their distances goes to zero. In certain situations treated by asymptotic analysis (e.g., homogenization) it can be shown that the solutions diverge with respect to the $W^{1,p}$ – norm for all $p > 2$.

Conclusion: Estimates using L_∞ norms of coefficients may be very coarse.

.....
There are two possible ways out:

1. Estimate $\|\nabla \tilde{u}\|$ using additional regularity of this function, which has derivatives integrable in L^p .
2. Use an approximation \tilde{v} (i.e., numerical solution of the simplified problem) instead of \tilde{u} .

.....
We consider the second way.

Problem \mathcal{P} .

$$\begin{aligned}\operatorname{div} A \nabla \tilde{u} + \tilde{f} &= 0 && \text{in } \Omega \\ u &= \tilde{u}_0 && \text{on } \Gamma_1, \\ A \nabla u \cdot \mathbf{n} &= F && \text{on } \Gamma_2\end{aligned}$$

Problem $\tilde{\mathcal{P}}$.

$$\begin{aligned}\operatorname{div} \tilde{A} \nabla \tilde{u} + \tilde{f} &= 0 && \text{in } \Omega \\ \tilde{u} &= u_0 && \text{on } \Gamma_1, \\ \tilde{A} \nabla \tilde{u} \cdot \mathbf{n} &= \tilde{F} && \text{on } \Gamma_2\end{aligned}$$

with simpler \tilde{A} , \tilde{f} , and \tilde{F} .

We will measure the simplification error in terms of the combined error norm

$$\|(u - \tilde{u}, p^* - \tilde{p}^*)\|_{V \times Y^*} := (\|(u - \tilde{u})\|_A^2 + \|p^* - \tilde{p}^*\|_{A^{-1}}^2)^{1/2}$$

It is convenient to split the error estimation problem into two parts using an *intermediate problem* $\tilde{\mathcal{P}}_+$:

$$\begin{aligned} \operatorname{div} A \nabla \tilde{u}_+ + \tilde{f} &= 0 && \text{in } \Omega, \\ \tilde{u}_+ &= u_0 && \text{on } \Gamma_1, \\ A \nabla \tilde{u}_+ \cdot \mathbf{n} &= \tilde{F} && \text{on } \Gamma_2. \end{aligned}$$

Then the total error is split

$$\begin{aligned} \|(u - \tilde{u}, p^* - \tilde{p}^*)\|_{V \times Y^*} \\ \leq \|(u - \tilde{u}_+, p^* - \tilde{p}_+^*)\|_{V \times Y^*} + \|(\tilde{u}_+ - \tilde{u}, \tilde{p}_+^* - \tilde{p}^*)\|_{V \times Y^*} \quad (1) \end{aligned}$$

Assume that the "simplified" functions \tilde{f} and \tilde{F} satisfy the conditions

$$\{f - \tilde{f}\}_{\Omega_i} = 0 \text{ and } \{F - \tilde{F}\}_{\Gamma_k} = 0, \quad i = 1, 2, \dots, N; \quad k = 1, 2, \dots, K,$$

where Ω_i is a collection of non-overlapping sets such that $\overline{\Omega} = \cup_{i=1}^N \overline{\Omega}_i$ and $\Gamma_k, k = 1, 2, \dots, K$ is a non-overlapping covering of the boundary Γ_2 .

Estimation of

$$\|(u - \tilde{u}_t, p^* - \tilde{p}_t^*)\|_{V \times Y^*}$$

We can use the estimate

$$\|\nabla(u - \tilde{u}_t)\|_A \leq \|A\nabla\tilde{u}_t - \mathbf{y}^*\|_{A^{-1}} + \left(\sum_{i=1}^N C_{1i}^2 \|\mathcal{R}(\mathbf{y}^*)\|_{\Omega_i}^2 \right)^{1/2} + \left(\sum_{k=1}^K C_{2k}^2 \|\mathbf{y}^* \cdot \mathbf{n} - F\|_{\Gamma_{2k}}^2 \right)^{1/2}. \quad (2)$$

where $\mathcal{R}(\mathbf{y}^*) = \operatorname{div} \mathbf{y}^* + f$, and $C_{1i}(\Omega_i)$ and $C_{2k}(\Gamma_k)$ are constants in the Poincare type inequalities

$$\inf_{c \in \mathbb{R}} \|w - c\|_{\Omega} \leq C_1 \|\nabla w\|_{\Omega}, \quad \inf_{c \in \mathbb{R}} \|w - c\|_{\Gamma} \leq C_1 \|\nabla w\|_{\Omega}$$

Computable bounds of these constants are known. \square

Set $\mathbf{y}^* = A \nabla \tilde{u}_+$. Then

$$\|\nabla(u - \tilde{u}_+)\|_A \leq \epsilon(\tilde{f}, \tilde{F}) := \left(\sum_{i=1}^N C_{1i}^2 \|f - \tilde{f}\|_{\Omega_i}^2 \right)^{1/2} + \left(\sum_{k=1}^K \tilde{C}_{2k}^2 \|F - \tilde{F}\|_{\Gamma_k}^2 \right)^{1/2},$$

Since $\mathbf{p}^* = A \nabla u$ and $\tilde{\mathbf{p}}_+^* = A \nabla \tilde{u}_+$, the estimate (2) yields

$$\|\mathbf{p}^* - \tilde{\mathbf{p}}_+^*\|_{A^{-1}} \leq \epsilon(\tilde{f}, \tilde{F})$$

and we conclude that

$$\|(u - \tilde{u}_+)\|_A^2 + \|\mathbf{p}^* - \tilde{\mathbf{p}}_+^*\|_{A^{-1}}^2 \leq 2\epsilon^2(\tilde{f}, \tilde{F}).$$

The quantity $\epsilon^2(\tilde{f}, \tilde{F})$ is easily computable.

Moreover, it can be used to find "optimal" splittings of Ω and Γ that minimise the corresponding error.

Estimation of

$$\|(\tilde{u}_+ - \tilde{u}, \tilde{\mathbf{p}}_+^* - \tilde{\mathbf{p}}^*)\|_{V \times Y^*}$$

This error arises exclusively due to the difference between \tilde{A} and A . Now we consider $\tilde{\mathcal{P}}_+$ as the "original" problem and $\tilde{\mathcal{P}}$ as its simplification. We have the estimate analogous to (2):

$$\|\nabla(\tilde{u}_+ - \tilde{u})\|_A \leq \|A\nabla\tilde{u} - \mathbf{y}^*\|_{A^{-1}} + \left(\sum_{i=1}^N C_{1i}^2 \|\mathcal{R}(\mathbf{y}^*)\|_{\Omega_i}^2 \right)^{1/2} + \left(\sum_{k=1}^K C_{2k}^2 \|\mathbf{y}^* \cdot \mathbf{n} - F\|_{\Gamma_{2k}}^2 \right)^{1/2}. \quad (3)$$

Since solutions of $\tilde{\mathcal{P}}$ satisfy $\mathcal{R}(\tilde{\mathbf{p}}^*) = \operatorname{div} \tilde{\mathbf{p}}^* + \tilde{f} = 0$, $\tilde{\mathbf{p}}^* \cdot \mathbf{n} = F$ on Γ_2 , the last two terms in (3) vanish, but now the first term is positive

$$\|A\nabla\tilde{u} - \mathbf{y}^*\|_{A^{-1}} = \|(A - \tilde{A})\nabla\tilde{u}\|_{A^{-1}}$$

We arrive at the estimate

$$\|(\boldsymbol{u} - \tilde{\boldsymbol{u}}_+)\|_A^2 + \|\boldsymbol{p}^* - \tilde{\boldsymbol{p}}_+^*\|_{A^{-1}}^2 = \int_{\Omega} \boldsymbol{D} \nabla \tilde{\boldsymbol{u}} \cdot \nabla \tilde{\boldsymbol{u}} \, d\boldsymbol{x}$$

where

$$\boldsymbol{D} := (\boldsymbol{A} - \tilde{\boldsymbol{A}}) \boldsymbol{A}^{-1} (\boldsymbol{A} - \tilde{\boldsymbol{A}})$$

is the **defect matrix** that reflects the difference between \boldsymbol{A} and $\tilde{\boldsymbol{A}}$ (\boldsymbol{D} is positive semidefinite).

Joining (2) and (3), we obtain

$$\boldsymbol{e}_{\text{mod}} := \|(\boldsymbol{u} - \tilde{\boldsymbol{u}}, \boldsymbol{p}^* - \tilde{\boldsymbol{p}}^*)\|_{\boldsymbol{V} \times \boldsymbol{Y}^*} \leq \|\nabla \tilde{\boldsymbol{u}}\|_{\boldsymbol{D}} + \sqrt{2}\epsilon(\tilde{\boldsymbol{f}}, \tilde{\boldsymbol{F}}).$$

A lower bound is derived analogously

$$\|(u - \tilde{u}, p^* - \tilde{p}^*)\|_{V \times Y^*} \geq \|\nabla \tilde{u}\|_D - \sqrt{2}\epsilon(\tilde{f}, \tilde{F}).$$

In general $\|\nabla \tilde{u}\|_D$ is unknown

To make the estimates fully computable, we use a suitable approximation $\tilde{v} \in V_0 + u_0$ of \tilde{u} .

$$e_{\text{mod}} \leq \mathcal{E}(\tilde{v}) + \sqrt{2}\epsilon(\tilde{f}, \tilde{F}), \quad \mathcal{E}(\tilde{v}) := \|\nabla \tilde{v}\|_D + \kappa_D \|\nabla \tilde{e}\|$$

where $\kappa_D = \text{ess sup}_{\mathbf{x} \in \Omega} |D(\mathbf{x})|$ and $\tilde{e} = \tilde{u} - \tilde{v}$.

The approximation error $\|\nabla \tilde{e}\| = \|\nabla(\tilde{u} - \tilde{v})\|$ can be estimated by the same technology. Finally, we get the estimate

$$\begin{aligned} e_{\text{mod}} \leq \|\nabla \tilde{v}\|_D + \kappa_D \Big(& \|\tilde{A} \nabla \tilde{v} - \tilde{\mathbf{y}}^*\|_{\tilde{A}^{-1}} + C_1 \|\mathcal{R}(\tilde{\mathbf{y}}^*)\|_{\Omega} \\ & + C_2 \|\tilde{\mathbf{y}}^* \cdot \mathbf{n} - F\|_{\Gamma_2} \Big) + \sqrt{2}\epsilon(\tilde{f}, \tilde{F}) \end{aligned}$$

that contains approximations \tilde{v} and $\tilde{\mathbf{y}}^*$ of the simplified problem.

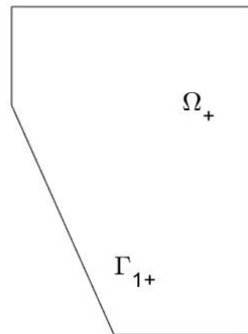
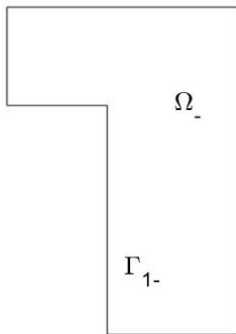
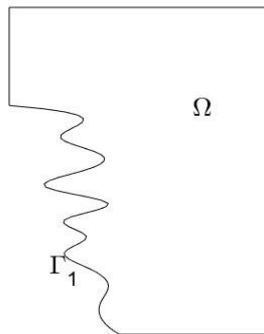
$$G(\Lambda v) + F(v) \Leftrightarrow \tilde{G}(\Lambda v) + \tilde{F}(v)$$

If the simplified problem is generated by the functionals $\tilde{\mathcal{G}}$ and $\tilde{\mathcal{F}}$, then the corresponding modeling error satisfies the identity

$$\mu(\tilde{u}) + \mu(\tilde{p}^*) = \mathcal{D}_{\mathcal{G}}(\Lambda \tilde{u}, \tilde{p}^*) + \mathcal{D}_{\mathcal{F}}(\tilde{u}, -\Lambda^* \tilde{p}^*).$$

The right hand side of this identity contains solutions of the simplified problem and "complicated" coefficients of the original problem enter the integrals associated with $\mathcal{D}_{\mathcal{G}}$ and $\mathcal{D}_{\mathcal{F}}$.

Simplification of the Dirichlet boundary



$$\Omega_- \subset \Omega \subset \Omega_+$$

where Ω_- and Ω_+ are two "simple" domains with Lipschitz boundaries Γ_- and Γ_+ .

Problem \mathcal{P} : minimisation of

$$\mathcal{J}_\Omega(v) := \frac{1}{2} \int_{\Omega} (A \nabla v \cdot \nabla v - f v) d\mathbf{x}$$

over the set $V(\Omega) := \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_1\}$.

$$\boxed{\operatorname{div} A \nabla u + f = 0}$$

Now A and f are "simple", but the domain is "complicated".

.....

Comment: we assume (for simplicity) that on Γ_2 zero Neumann condition is imposed, but this is not essential and non-homogeneous conditions can be considered analogously

We consider two variational problems generated by the functionals

$$\mathcal{J}_{\Omega_+}(v_+) := \frac{1}{2} \int_{\Omega_+} (\hat{A} \nabla \hat{v}_+ \cdot \nabla v_+ - \hat{f} \hat{v}_+) d\mathbf{x}$$

$$\mathcal{J}_{\Omega_-}(v_-) := \frac{1}{2} \int_{\Omega_-} (A \nabla v_- \cdot \nabla v_- - f v_-) d\mathbf{x}$$

which are minimized in the sets

$$V(\Omega_+) := \{v_+ \in H^1(\Omega_+) \mid v_+ = 0 \text{ on } \Gamma_{1+}\}$$

$$V(\Omega_-) := \{v \in H^1(\Omega_-) \mid v_- = 0 \text{ on } \Gamma_{1-}\},$$

Extensions:

Let \hat{f} be the extension of f to Ω_+ by zero, so that $\hat{f} \in L^2(\Omega_+)$.

If A does not depend on \mathbf{x} , we set $\hat{A} = A$.¹

By \hat{v} we denote extension of v defined in Ω or Ω_- to Ω_+ .

¹In general, $\hat{A}(\mathbf{x}) = A(\mathbf{x})$ if $\mathbf{x} \in \Omega$ and $\tilde{A}(\mathbf{x}) = A_+$ in $\omega := \Omega_+ \setminus \Omega$, where A_+ is a positive definite extension, whose eigenvalues lie between λ_{\ominus} and λ_{\oplus} .

Analysis is based on the identity

$$\frac{1}{2} \|\nabla(u_+ - \hat{u})\|_{\hat{A}, \Omega_+}^2 = \mathcal{J}_{\Omega_+}(\hat{u}) - \mathcal{J}_{\Omega_+}(u_+) \quad (4)$$

which holds because $u_+ \in V_+(\Omega_+)$ is the minimizer, and $\hat{u} \in V_+(\Omega_+)$. Next, $u \in V(\Omega)$ minimizes the functional \mathcal{J}_Ω and for any function $v_- \in V(\Omega_-)$ we have

$$\mathcal{J}_{\Omega_+}(\hat{u}) = \mathcal{J}_\Omega(u) \leq \mathcal{J}_\Omega(\hat{v}_-) = \mathcal{J}_{\Omega_+}(\hat{v}_-).$$

Hence

$$\frac{1}{2} \|\nabla(u_+ - \hat{u})\|_{\hat{A}, \Omega_+}^2 \leq \mathcal{J}_{\Omega_+}(\hat{v}_-) - \mathcal{J}_{\Omega_+}(u_+) = \frac{1}{2} \|\nabla(u_+ - \hat{v}_-)\|_{\hat{A}, \Omega_+}^2,$$

and we conclude that

$$\|\nabla(u_+ - \hat{u})\|_{\hat{A}, \Omega_+} \leq \|\nabla(u_+ - \hat{v}_-)\|_{\hat{A}, \Omega_+}. \quad (5)$$

Let $\widehat{V}(\Omega_-)$ denote the space containing extensions by zero of the functions in $V(\Omega_-)$ and $\omega := \Omega_+ \setminus \Omega$. Using (5), we obtain

$$\begin{aligned} \|\nabla u_+\|_{\widehat{A},\omega}^2 + \|\nabla(u_+ - \widehat{u})\|_{A,\Omega}^2 &= \|\nabla(u_+ - \widehat{u})\|_{\widehat{A},\Omega_+}^2 \leq \|\nabla(u_+ - \widehat{v}_-)\|_{\widehat{A},\Omega_+}^2 \\ &= \|\nabla u_+\|_{\widehat{A},\omega}^2 + \|\nabla(u_+ - \widehat{v}_-)\|_{A,\Omega}^2 \quad \forall \widehat{v}_- \in \widehat{V}(\Omega_-). \end{aligned}$$

Hence for u_+ we have the estimate of $e_{\text{mod}}(\Omega_+)$:

$$\|\nabla(u_+ - u)\|_{A,\Omega} \leq \inf_{\widehat{v}_- \in \widehat{V}(\Omega_-)} \|\nabla(u_+ - \widehat{v}_-)\|_{A,\Omega} =: \Pi_{\Omega_-}(u_+).$$

This holds for any $\Omega_- \subset \Omega$, so in particular we can take $\Omega_- = \Omega$. Then the equality holds:

$$e_{\text{mod}}(\Omega_+) = \|\nabla(u_+ - u)\|_{A,\Omega} = \inf_{\widehat{v} \in \widehat{V}(\Omega)} \|\nabla(u_+ - \widehat{v})\|_{A,\Omega}.$$

However, u_+ is unknown!

We need a bound that operates with approximate solutions only. It has the form

$$\|\nabla(u_+ - u)\|_{A, \Omega} \leq \|A \nabla v_+ - \mathbf{y}_+^*\|_{A^{-1}, \Omega_+} + C_{\Omega_+} \|\operatorname{div} \mathbf{y}_+^* + f\|_{\Omega_+} + \Pi_{\Omega_-}(v_+)$$

Now the term related to boundary simplification includes only known functions

$$\Pi_{\Omega_-}(v_+) = \inf_{\hat{v}_- \in \hat{V}(\Omega_-)} \|\nabla(v_+ - \hat{v}_-)\|_{A, \Omega}.$$

Here $C_{\Omega_+} = \frac{C_F(\Omega_+)}{\lambda_{\ominus}^{1/2}(A)}$, i.e., the constant is related to **simple** Ω_+ .

v_+ and \mathbf{y}_+^* are numerical approximations of the exact solutions u_+ and \mathbf{p}_+^* of the simplified problem.

Notice that finding an upper bound of $\Pi_{\Omega_-}(v_+)$ can be found by solving finite dimensional problem associated with "simple" domains Ω_- and Ω_+ only.

Generalization to the whole class $J(v) = G(\Lambda v) + F(v)$

$$\mathcal{D}_{\mathcal{G}}(\Lambda u, p_+^*) \leq \inf_{v_- \in \widehat{V}(\Omega_-)} \mathcal{D}_{\mathcal{G}}(\Lambda \widehat{v}_-, p_+^*).$$

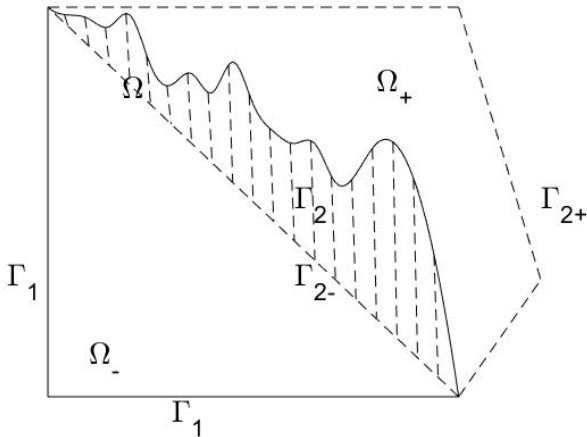
The left hand side is the measure of the distance between exact solutions of the original and the simplified problem associated with the domain Ω_+ .

.....
Recall that $p_+^* = p^*(\Lambda u_+)$, e.g., if \mathcal{G} is differentiable, then $p_+^* = \mathcal{G}'(\Lambda u_+)$. Therefore, $\mathcal{D}_{\mathcal{G}}(\Lambda u, p_+^*)$ can be viewed as a measure between Λu and Λu_+ , where u_+ is understood as the restriction of the function u to Ω .

For example, if G is generated by α -Laplacian, then

$$\mathcal{D}_{\mathcal{G}}(\Lambda u, p_+^*) = \int_{\Omega} \left(\frac{1}{\alpha} |\nabla u|^\alpha + \frac{1}{\alpha^*} |\nabla u_+|^\alpha - \nabla u \cdot \nabla u_+ |\nabla u_+|^{\alpha-2} \right) d\mathbf{x}.$$

Simplification of the Neumann boundary



see the book

Dimension reduction

Systematic study of mathematical models associated with the reduction of dimensions has began in the 19th century mainly due to the development of solid mechanics ([Kirchhoff](#), [Love](#), [Timoshenko](#), [Reissner](#), [Mindlin](#)...)

Analysis of asymptotic convergence:

- A. L. Alessandrini, D. N. Arnold, R. S. Falk, A. L. Madureira, *Derivation and justification of plate models by variational methods*, in *Plates and Shells*, Quebec 1996,
- D. Morgenstern, *Herleitung der Plattentheorie aus der dreidimensionalen Elastizitätstheorie*, Arch. Rational. Mech. Anal. 4, 145-152, (1959).
- G. Anzellotti, S. Baldo, D. Percivale, *Dimension reduction in variational problems, asymptotic development in Γ -convergence and thin structures in elasticity*, Asymptotic Analysis, vol. 9, no. 1, 61-100, (1994).
- I. Babuška, I. Lee, C. Schwab, *On the a posteriori estimation of the modeling error for the heat conduction in a plate and its use for adaptive hierarchical modeling*, in *Proceedings of the Third ARO Workshop on Adaptive Methods for Partial Differential Equations* (Troy, NY, 1992), volume 14, 5–21, (1994).

- B. A. Shoikhet, *On asymptotically exact equations of thin plates of complex structure*, J. Appl. Math. Mech. 37 (1973), 867–877, (1974).
- D. Braess, S. Sauter, C. Schwab, *On the justification of plate models*, J. Elasticity, 103, 1, 53–71, (2011).
- C. Schwab, *A-posteriori modeling error estimation for hierarchic plate models*, Numer. Math., 74, 221–259, (1996).
- P. G. Ciarlet, P. Destuynder, *A justification of a nonlinear model in plate theory*, Comput. Methods Appl. Mech. Engrg. 17/18, 227–258 (1979).
- B. Miara, *Justification of the asymptotic analysis of elastic plates, I. The linear case*, Asymptotic Analysis, 9(1), 47–60, (1994).

General scheme of dimension reduction

Original
Problem

\mathcal{P}

u, p^*

\Rightarrow

Reduced
Problem

$\hat{\mathcal{P}}$

\hat{u}, \hat{p}^*

\Rightarrow

Computable(Discrete)
Problem

$\hat{\mathcal{P}}_{\text{com}}$

\hat{v}, \hat{q}^*

\downarrow

$\mathcal{R}\hat{v}, \mathcal{R}^*\hat{q}^*$

In general, the functions \hat{v} and \hat{q}^* do not belong to V and Y^* , respectively. For this reason any comparison with the exact solutions should use specially constructed *reconstruction operators*

$$\mathfrak{R} : \hat{V} \rightarrow V \quad \text{and} \quad \mathfrak{R}^* : \hat{Y}^* \rightarrow Y^*.$$

The functions

$$u_{\mathfrak{R}} := \mathfrak{R}\hat{u} \in V \quad \text{and} \quad p_{\mathfrak{R}}^* := \mathfrak{R}^*\hat{p}^* \in Y^*$$

are considered as reconstructions of the exact solutions obtained by the dimension reduction method.

Dimension reduction errors are formed by reconstructions

$$\|u - u_{\mathfrak{R}}\|_V$$

$$\|p^* - p_{\mathfrak{R}}^*\|_{Y^*}$$

$$\|u - u_{\mathfrak{R}}\|_V^2 + \|p^* - p_{\mathfrak{R}}^*\|_{Y^*}^2$$

Reconstruction operator \mathfrak{R} may be different for different classes of problems, but it must satisfy two principal conditions:

computational simplicity and boundedness.

However, weed more, e.g., for linear problems, it suffices to assume that \mathfrak{R} (and/or \mathfrak{R}^*) additionally satisfies the Lipschitz condition

$$\|\mathfrak{R}\hat{v}_1 - \mathfrak{R}\hat{v}_2\|_V \leq C_{\mathfrak{R}} \|\hat{v}_1 - \hat{v}_2\|_{\hat{V}} \quad \forall \hat{v}_1, \hat{v}_2 \in \hat{V},$$

where $C_{\mathfrak{R}} > 0$ is known and does not depend on \hat{v}_1 and \hat{v}_2 .

In practice, we usually know only an approximate solution \hat{v} and the respective reconstruction $\mathcal{R}\hat{v}$ is what we indeed have.

The quantity

$$e_{\text{com}} := \|\hat{u} - \hat{v}\|_{\hat{v}}$$

is the error arising when a differential problem is replaced by a computable (finite dimensional) counterpart. By the Lipschitz condition, we find that

$$\begin{aligned} \|u - \mathcal{R}\hat{v}\|_V &\leq \|u - \mathcal{R}\hat{u}\|_V + \|\mathcal{R}\hat{u} - \mathcal{R}\hat{v}\|_V \\ &\leq \|u - \mathcal{R}\hat{u}\|_V + C_{\mathcal{R}} \|\hat{u} - \hat{v}\|_{\hat{v}} = e_{\text{mod}} + C_{\mathcal{R}} e_{\text{com}}, \end{aligned} \quad (6)$$

where

$$e_{\text{mod}} := \|u - \mathcal{R}\hat{u}\|_V$$

is the modelling error.

Second order elliptic problems

- S. R., S. Sauter, A. Smolianski, *A posteriori estimation of dimension reduction errors for elliptic problems on thin domains*, SIAM J. Num. Anal., 42(4), 1435–1451 (2004).

We consider 3D domains $\Omega = \hat{\Omega} \times (t_{\ominus}(x_1, x_2), t_{\oplus}(x_1, x_2))$

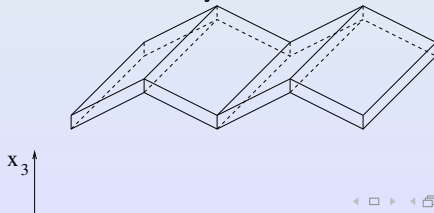
t_{\ominus} and t_{\oplus} are Lipschitz continuous functions defined in $\hat{\Omega}$ and $\hat{\Omega}$ is a bounded domain with Lipschitz boundary $\hat{\Gamma}$. The sets

$$\Gamma_{\ominus} := \{x \in \mathbb{R}^3 \mid (x_1, x_2) \in \hat{\Omega}, x_3 = t_{\ominus}(x_1, x_2)\},$$

$$\Gamma_{\oplus} := \{x \in \mathbb{R}^3 \mid (x_1, x_2) \in \hat{\Omega}, x_3 = t_{\oplus}(x_1, x_2)\},$$

$$\Gamma_0 := \{x \in \mathbb{R}^3 \mid (x_1, x_2) \in \hat{\Gamma}, t_{\ominus}(x_1, x_2) < x_3 < t_{\oplus}(x_1, x_2)\}$$

denote different parts of the boundary Γ



Ω is a “thin” domain if there exists a ball $\widehat{B}_R \subset \widehat{\Omega}$ such that

$$R \gg \sup_{(x_1, x_2) \in \widehat{\Omega}} t(x_1, x_2)$$

where $t = t_{\oplus} - t_{\ominus}$ is the “thickness” function.

In general, it is not required that t is constant but we assume that

$$t(x_1, x_2) \geq t_* > 0 \quad \forall (x_1, x_2) \in \overline{\widehat{\Omega}}.$$

In Ω , we consider the problem (also called *Problem \mathcal{P}*)

$$-\operatorname{div}(A\nabla u) = f \quad \text{in } \Omega, \quad (7)$$

$$u = 0 \quad \text{on } \Gamma_0, \quad (8)$$

$$A\nabla u \cdot \mathbf{n}_\ominus = F_\ominus \quad \text{on } \Gamma_\ominus, \quad (9)$$

$$A\nabla u \cdot \mathbf{n}_\oplus = F_\oplus \quad \text{on } \Gamma_\oplus, \quad (10)$$

where $f \in L_2(\Omega)$, $F_\ominus \in L_2(\Gamma_\ominus)$, $F_\oplus \in L_2(\Gamma_\oplus)$, \mathbf{n}_\ominus and \mathbf{n}_\oplus are two outward normal vectors associated with Γ_\ominus and Γ_\oplus respectively, and $A \in L^\infty(\mathbb{M}_s^{3 \times 3})$ is a uniformly positive definite matrix:

$$\lambda_\ominus |\zeta|^2 \leq A(x)\zeta \cdot \zeta \leq \lambda_\oplus |\zeta|^2 \quad \forall \zeta \in \mathbb{R}^3 \text{ a. e. in } \Omega. \quad (11)$$

From now on we set $\hat{\mathbf{x}} = (x_1, x_2) \in \hat{\Omega}$, mark all functions depending only on (x_1, x_2) by $\hat{}$, and use different notation for 3- and 2-dimensional operators, e.g.,

$$\operatorname{div} \mathbf{q} = \frac{\partial q_1}{\partial x_1} + \frac{\partial q_2}{\partial x_2} + \frac{\partial q_3}{\partial x_3}, \quad \widehat{\operatorname{div}} \hat{\mathbf{q}} = \frac{\partial \hat{q}_1}{\partial x_1} + \frac{\partial \hat{q}_2}{\partial x_2}.$$

Also, we use the notation

$$\hat{\mathbf{F}}_{\ominus}(\hat{\mathbf{x}}) := F_{\ominus}(\hat{\mathbf{x}}, t_{\ominus}(\hat{\mathbf{x}})), \quad \hat{\mathbf{F}}_{\oplus}(\hat{\mathbf{x}}) := F_{\oplus}(\hat{\mathbf{x}}, t_{\oplus}(\hat{\mathbf{x}})), \quad \hat{\mathbf{x}} \in \hat{\Omega}.$$

The generalized solution $u \in V_0$ satisfies

$$\int_{\Omega} A \nabla u \cdot \nabla w \, d\mathbf{x} = \int_{\Omega} f w \, d\mathbf{x} + \int_{\Gamma_{\ominus}} F_{\ominus} w \, ds + \int_{\Gamma_{\oplus}} F_{\oplus} w \, ds \quad \forall w \in V_0, \quad (12)$$

where $V_0 := \{v \in H^1(\Omega) \mid v = 0 \text{ on } \Gamma_0\}$.

Zero-order reduced model is based on the following hypothesis:

the exact solution u is almost constant with respect to x_3 .

In this model, it is natural to define the reconstruction operator \mathfrak{R} by the relation

$$v(\hat{\mathbf{x}}, x_3) := \hat{v}(\hat{\mathbf{x}}) \quad \text{for any } t_{\ominus}(\hat{\mathbf{x}}) \leq x_3 \leq t_{\oplus}(\hat{\mathbf{x}}), \quad (\hat{\mathbf{x}}, x_3) \in \Omega.$$

In other words, for $\hat{v} \in \mathring{H}^1(\hat{\Omega})$ the recovered 3D function $\mathfrak{R}\hat{v} \in V_0^+(\Omega) \subset V_0(\Omega)$ is defined as the 3D-function independent of x_3 and

$$\text{Im}\mathfrak{R} = V_0^+(\Omega) := \{v \in V_0(\Omega) \mid \exists \hat{v} \in \hat{V}_0(\hat{\Omega}), v(\hat{\mathbf{x}}, x_3) = \hat{v}(\hat{\mathbf{x}})\}.$$

Warning: above made assumptions serve only as an intuitive motivation for the introduction of the subspace V_0^+ . In a particular case related to a particular problem, it is not guaranteed a priori that the solution in V_0^+ serves as a good approximation of u .

Energy-norm projection $u_{\mathfrak{R}}$ of the exact solution u onto the reduced subspace $V_0^+(\Omega)$ leads to the problem

$$\int_{\Omega} A \nabla u_{\mathfrak{R}} \cdot \nabla w \, d\mathbf{x} = \int_{\Omega} f w \, d\mathbf{x} + \int_{\Gamma_{\ominus}} F_{\ominus} w \, ds + \int_{\Gamma_{\oplus}} F_{\oplus} w \, ds \quad \forall w \in V_0^+ \quad (13)$$

Since $w_{,3} := \frac{\partial w}{\partial x_3} = 0$ we observe that

$$\int_{\Omega} A \nabla u_{\mathfrak{R}} \cdot \nabla w \, d\mathbf{x} = \int_{\hat{\Omega}} t(\hat{\mathbf{x}}) \tilde{A}_p \hat{\nabla} \hat{u} \cdot \hat{\nabla} \hat{w} \, d\hat{\mathbf{x}},$$

where \hat{u} is the plane part of the function $u_{\mathfrak{R}}$, $\hat{\nabla} \hat{w} = (\hat{w}_{,1}, \hat{w}_{,2})$ denotes the plane gradient, $A_p(x) = (a_{ij}(x))$, $i, j = \{1, 2\}$ is the plane part of the matrix A , and $\tilde{A}_p(\hat{\mathbf{x}}) = (\tilde{a}_{ij}(\hat{\mathbf{x}}))$ is the plane part of A averaged with respect to x_3 and tilde denotes the x_3 -averaging, i.e.,

$$\tilde{g}(\hat{\mathbf{x}}) := \frac{1}{t(\hat{\mathbf{x}})} \int_{t_{\ominus}(\hat{\mathbf{x}})}^{t_{\oplus}(\hat{\mathbf{x}})} g(\hat{\mathbf{x}}, x_3) \, dx_3 \quad \text{for any } g \in L_1(\Omega).$$

Now we rearrange the terms in the right hand side of (13)

$$\begin{aligned}\int_{\Gamma_{\ominus}} F_{\ominus} \widehat{w} \, ds &= \int_{\widehat{\Omega}} \widehat{F}_{\ominus}(\widehat{\mathbf{x}}) \widehat{w}(\widehat{\mathbf{x}}) \chi_{\ominus}(\widehat{\mathbf{x}}) \, d\widehat{\mathbf{x}}, \\ \int_{\Gamma_{\oplus}} F_{\oplus} \widehat{w} \, ds &= \int_{\widehat{\Omega}} \widehat{F}_{\oplus}(\widehat{\mathbf{x}}) \widehat{w}(\widehat{\mathbf{x}}) \chi_{\oplus}(\widehat{\mathbf{x}}) \, d\widehat{\mathbf{x}},\end{aligned}$$

where²

$$\chi_{\oplus}(\widehat{\mathbf{x}}) := \sqrt{1 + |\widehat{\nabla} t_{\oplus}(\widehat{\mathbf{x}})|^2} \quad \text{and} \quad \chi_{\ominus}(\widehat{\mathbf{x}}) := \sqrt{1 + |\widehat{\nabla} t_{\ominus}(\widehat{\mathbf{x}})|^2}.$$

Therefore, problem (13) is transformed into a *reduced Problem $\widehat{\mathcal{P}}$* :
Find $\widehat{\mathbf{u}} \in \widehat{V}_0(\widehat{\Omega})$ such that

$$\int_{\widehat{\Omega}} t(\widehat{\mathbf{x}}) \widetilde{A}_p(\widehat{\mathbf{x}}) \widehat{\nabla} \widehat{\mathbf{u}} \cdot \widehat{\nabla} \widehat{w} \, d\widehat{\mathbf{x}} = \int_{\widehat{\Omega}} t(\widehat{\mathbf{x}}) \widehat{f}(\widehat{\mathbf{x}}) \widehat{w} \, d\widehat{\mathbf{x}} \quad \forall \widehat{w} \in \widehat{V}_0. \quad (14)$$

²If the functions t_{\ominus} and t_{\oplus} are piecewise smooth, then these integrals should be presented as sums of the corresponding integrals associated with subdomains where their gradients exist.

In (14), the source term is defined by the relation

$$\hat{f}(\hat{\mathbf{x}}) = \tilde{f}(\hat{\mathbf{x}}) + \frac{\hat{F}_{\ominus}(\hat{\mathbf{x}})\chi_{\ominus}(\hat{\mathbf{x}}) + \hat{F}_{\oplus}(\hat{\mathbf{x}})\chi_{\oplus}(\hat{\mathbf{x}})}{t(\hat{\mathbf{x}})}.$$

(*Problem* $\hat{\mathcal{P}}$) is a two-dimensional elliptic problem

$$\begin{aligned} -\widehat{\operatorname{div}}(t(\hat{\mathbf{x}})\tilde{A}_p(\hat{\mathbf{x}})\widehat{\nabla}\hat{u}) &= t(\hat{\mathbf{x}})\hat{f}(\hat{\mathbf{x}}) \quad \text{in } \hat{\Omega} \\ \hat{u} &= 0 \quad \text{on } \hat{\Gamma}. \end{aligned} \tag{15}$$

Then the recovered function is defined by the relation $u_{\mathfrak{R}} = \mathfrak{R}\hat{u}$ and the error of dimension reduction is

$$\mathbf{e} := \mathbf{u} - u_{\mathfrak{R}} = \mathbf{u} - \mathfrak{R}\hat{u}.$$

How to get a guaranteed bound of the modeling error?

- take the deviation estimate for 3D model
- substitute there reconstructions of 2D solutions (first error bound)
- insert some (simple) correction terms to have a better reconstruction and better estimate

Estimate for the 3D diffusion problem ($e = u - u_{\mathfrak{R}}$):

$$\|\nabla e\|_A^2 \leq (1 + \gamma) \left(M_1^2(u_{\mathfrak{R}}, \mathbf{y}^*) + \frac{1 + \delta}{\gamma} C_1^2 M_2^2(\mathbf{y}^*) + \frac{1 + \delta}{\gamma \delta} C_2^2 M_3^2(\mathbf{y}^*) \right), \quad (16)$$

where γ and δ are arbitrary positive numbers,

$$\mathbf{y}^* \in Q_{\Lambda^*}^* := \left\{ \mathbf{y}^* \in L_2(\Omega, \mathbb{R}^3) \mid \operatorname{div} \mathbf{y}^* \in L_2(\Omega), \right. \\ \left. \mathbf{y}^* \cdot \mathbf{n}_{\ominus} \in L_2(\Gamma_{\ominus}), \mathbf{y}^* \cdot \mathbf{n}_{\oplus} \in L_2(\Gamma_{\oplus}) \right\},$$

$$M_1^2(u_{\mathfrak{R}}, \mathbf{y}^*) := \int_{\Omega} (\nabla u_{\mathfrak{R}} - A^{-1} \mathbf{y}^*) \cdot (A \nabla u_{\mathfrak{R}} - \mathbf{y}^*) \, d\mathbf{x}, \\ M_2^2(\mathbf{y}^*) := \|\operatorname{div} \mathbf{y}^* + f\|_{\Omega}^2, \\ M_3^2(\mathbf{y}^*) := \|F_{\ominus} - \mathbf{y}^* \cdot \mathbf{n}_{\ominus}\|_{\Gamma_{\ominus}}^2 + \|F_{\oplus} - \mathbf{y}^* \cdot \mathbf{n}_{\oplus}\|_{\Gamma_{\oplus}}^2.$$

Now we need to find a suitable reconstruction to put instead of \mathbf{y}^*

A freedom of choosing γ and δ and the function \mathbf{y}^* can be used to make the estimate as sharp as possible. Certainly, the best possible choice for \mathbf{y}^* would be

$$\mathbf{p}^* = A \nabla u$$

In this case, $M_2 = M_3 = 0$ and $M_1 = \text{exact error}$.

However, u is unknown and instead of \mathbf{p}^* we have to choose a certain reconstruction $\mathbf{p}_{\mathfrak{R}}^*$ using the solution of the reduced problem.

In particular, we may approximate it by $\mathbf{p}_{\mathfrak{R}}^* = \tilde{A}_p \nabla u_{\mathfrak{R}} + \boldsymbol{\tau}^*$ where

$\boldsymbol{\tau}^* = \{0, 0, \psi(\mathbf{x})\}^T$ is a *correction term*,

$\tilde{A}_p \in \mathbb{M}_s^{3 \times 3}$ has the same entries as \tilde{A}_p in the "plane" part and zero values in the third row and third column.

$\psi \in L_2(\Omega)$ is an auxiliary function, we assume that it possesses an additional regularity, namely, $\frac{\partial \psi}{\partial x_3} \in L_2(\Omega)$, $\psi \in L_2(\Gamma_{\ominus})$, and $\psi \in L_2(\Gamma_{\oplus})$

Why ψ is required?

Note that

$$\operatorname{div} \mathbf{p}_{\mathcal{R}}^* = \underbrace{\widehat{\operatorname{div}} \widetilde{A}_p \widehat{\nabla} \widehat{u}} + \frac{\partial \psi}{\partial x_3}.$$

Compare with

$$\operatorname{div} \mathbf{p}^* = \underbrace{p_{,1}^* + p_{,2}^*}_{\text{plane part}} + p_{,3}^*$$

For example, if we consider the Poisson equation (i.e. $A = \mathbb{1}$), then

$$\operatorname{div} \mathbf{p}^* = \Delta u = \widehat{\operatorname{div}} \widehat{\nabla} u(\widehat{\mathbf{x}}, x_3) + \frac{\partial u}{\partial x_3}.$$

Hence if $\widehat{u}(\widehat{\mathbf{x}})$ is a function used to approximate the plane part of the exact solution, then ψ should provide a good approximation of the derivative in x_3 -direction.

Now we skip 3 pages of computations and come to the final result

Now we skip 3 pages of computations and come to the final result

Notation:

$$\mathbf{B} := \mathbf{A}^{-1}, \quad \mathbf{B}_p := (b_{ij})_{i,j=1,2}, \quad \mathbf{b}_3 := (b_{31}, b_{32})^T.$$

Plane normals

$$\hat{\mathbf{n}}_{\ominus} = (v_1^{\ominus}, v_2^{\ominus}), \quad \hat{\mathbf{n}}_{\oplus} = (v_1^{\oplus}, v_2^{\oplus}), \quad v_3^{\ominus} = -\frac{1}{\chi_{\ominus}}, \quad v_3^{\oplus} = \frac{1}{\chi_{\oplus}}$$

are formed by the components of \mathbf{n}_{\ominus} and \mathbf{n}_{\oplus} .

$$M_1^2(u_{\mathfrak{R}}, \mathbf{p}_{\mathfrak{R}}^*) = \int_{\hat{\Omega}} t(\hat{\mathbf{x}}) (\tilde{B}_p \tilde{A}_p - \hat{\mathbb{1}}) \hat{\nabla} \hat{u} \cdot \tilde{A}_p \hat{\nabla} \hat{u} d\hat{\mathbf{x}} \\ + \int_{\Omega} (b_{33} \psi^2 + 2(\mathbf{b}_3 \cdot \tilde{A}_p \hat{\nabla} \hat{u}) \psi) d\mathbf{x},$$

where $\hat{\mathbb{1}} \in \mathbb{M}^{2 \times 2}$ is the unit matrix.

$$M_2^2(\mathbf{p}_{\mathfrak{R}}^*) = \left\| f - \tilde{f} - \frac{\hat{F}_{\ominus} \chi_{\ominus} + \hat{F}_{\oplus} \chi_{\oplus}}{t} - \frac{\hat{\nabla} t}{t} \cdot \tilde{A}_p \hat{\nabla} \hat{u} + \frac{\partial \psi}{\partial x_3} \right\|_{\Omega}^2.$$

$$M_3^2(\mathbf{p}_{\mathfrak{R}}^*) = \|F_{\ominus} - \tilde{A}_p \hat{\nabla} \hat{u} \cdot \hat{\mathbf{n}}_{\ominus} - \psi \nu_3^{\ominus}\|_{\Gamma_{\ominus}}^2 + \|F_{\oplus} - \tilde{A}_p \hat{\nabla} \hat{u} \cdot \hat{\mathbf{n}}_{\oplus} - \psi \nu_3^{\oplus}\|_{\Gamma_{\oplus}}^2$$

The term M_3 can be eliminated if we properly select ψ

Let us set

$$\psi_1(\mathbf{x}, x_3) = \hat{\alpha}(\hat{\mathbf{x}}) x_3 + \hat{\beta}(\hat{\mathbf{x}}),$$

where the functions $\hat{\alpha}$ and $\hat{\beta}$ are chosen such that

$$\psi_1(\hat{\mathbf{x}}, t_{\oplus}(\hat{\mathbf{x}})) \nu_3^{\oplus} = \hat{Y}_{\oplus} \quad \text{and} \quad \psi_1(\hat{\mathbf{x}}, t_{\ominus}(\hat{\mathbf{x}})) \nu_3^{\ominus} = \hat{Y}_{\ominus},$$

where Y_{\ominus} , Y_{\oplus} are given quantities.

Since the components ν_3^{\ominus} , ν_3^{\oplus} belong to $L_{\infty}(\hat{\Omega})$ and cannot vanish in $\hat{\Omega}$, the functions $\hat{\alpha}$ and $\hat{\beta}$ are uniquely defined by these conditions

$$\begin{aligned} \hat{\alpha} &= \frac{1}{t} \left(\frac{\hat{Y}_{\oplus}}{\nu_3^{\oplus}} - \frac{\hat{Y}_{\ominus}}{\nu_3^{\ominus}} \right), \\ \hat{\beta} &= \frac{1}{t} \left(\frac{\hat{Y}_{\ominus}}{\nu_3^{\ominus}} t_{\oplus} - \frac{\hat{Y}_{\oplus}}{\nu_3^{\oplus}} t_{\ominus} \right). \end{aligned}$$

Then

$$M_3(\mathbf{p}_{\mathfrak{R}}^*) = 0,$$

and we arrive at the estimate

$$\|e\|_A^2 \leq (1 + \gamma) M_1^2(u_{\mathfrak{R}}, \mathbf{p}_{\mathfrak{R}}^*) + \left(1 + \frac{1}{\gamma}\right) C_1^2 M_2^2(\mathbf{p}_{\mathfrak{R}}^*),$$

where γ is any positive number. Minimization of the right-hand side with respect to $\gamma > 0$ yields the estimate

$$\|e\|_A \leq M := M_1(u_{\mathfrak{R}}, \mathbf{p}_{\mathfrak{R}}^*) + C_1 M_2(\mathbf{p}_{\mathfrak{R}}^*)$$

Comment:

We can consider more general (with respect to x_3) functions, e.g.,

$$\psi_2(x) = \psi_1(x) + \hat{\eta}(\hat{\mathbf{x}})(x_3 - t_{\oplus}(\hat{\mathbf{x}}))(x_3 - t_{\ominus}(\hat{\mathbf{x}})),$$

where $\hat{\eta}$ is a function in $L_2(\hat{\Omega})$, also implies $M_3(\mathbf{p}_{\mathfrak{R}}^*) = 0$.

Assume that

$$t_{\ominus} = -\frac{t_0}{2}, \quad t_{\oplus} = \frac{t_0}{2}, \quad t_0 = \text{const} > 0. \quad (17)$$

and, in addition,

$$A = A(\hat{\mathbf{x}}), \quad a_{31} = a_{32} = 0.$$

Then

$$B = B(\hat{\mathbf{x}}), \quad B_p = A_p^{-1}, \quad b_{33} = a_{33}^{-1}, \quad b_{31} = b_{32} = 0.$$

Then $\nu_{\alpha}^{\ominus} = \nu_{\alpha}^{\oplus} = 0$ for $\alpha = 1, 2$, $\nu_3^{\ominus} = -1$, $\nu_3^{\oplus} = 1$ and the function ψ_1 takes the simplest form

$$\psi_1(x) = \frac{\hat{F}_{\oplus}(\hat{\mathbf{x}}) + \hat{F}_{\ominus}(\hat{\mathbf{x}})}{t_0} x_3 + \frac{\hat{F}_{\oplus}(\hat{\mathbf{x}}) - \hat{F}_{\ominus}(\hat{\mathbf{x}})}{2}$$

$$M_1^2 = \int_{\Omega} a_{33}^{-1} \psi_1^2 d\mathbf{x}, \quad M_2 = \|f - \tilde{f}\|_{\Omega}. \quad (18)$$

Our error estimate is reduced to

$$\|e\|_A \leq \sqrt{\frac{t_0}{3}} \left(\int_{\hat{\Omega}} a_{33}^{-1} (\hat{F}_{\oplus}^2 + \hat{F}_{\ominus}^2 - \hat{F}_{\oplus} \hat{F}_{\ominus}) d\hat{\mathbf{x}} \right)^{1/2} + C_1 \|f - \tilde{f}\|_{\Omega}. \quad (19)$$

Assume that f does not depend on x_3 and $\hat{F}_{\ominus} = 0$. Then the the second term vanishes and we obtain a simple estimate

$$\|e\|_A^2 \leq \frac{t_0}{3a_{33}} \int_{\hat{\Omega}} \hat{F}_{\oplus}^2 d\hat{\mathbf{x}}. \quad (20)$$

If $a_{33} = 1$ and $\hat{F}_{\oplus} = \hat{F}_{\ominus} = \hat{F}$, then the estimate

$$\|e\|_A \leq \sqrt{\frac{t_0}{3}} \|\hat{F}\|_{L_2(\hat{\Omega})} \quad (21)$$

is exactly the estimate deduced for the zero-order reduced model by I. Babuška, I. Lee, C. Schwab (1994).

Examples. We consider a two-dimensional test problem in the “sin-shape” domain depicted in Figure 1, whose upper and lower faces are given by the relation

$$t_{\oplus,\ominus}(x) = \sin(k\pi x) \pm \frac{t_0}{2}, \quad k = 1, 2, \dots,$$

where the constant $t_0 > 0$ is the domain thickness. In this example,

$$\hat{\Omega} = (0, 1) \quad \text{and} \quad \Omega = \{(x, y) \in \mathbb{R}^2 \mid x \in \hat{\Omega}, t_{\ominus}(x) < y < t_{\oplus}(x)\}.$$

The problem is

$$\begin{aligned} -\Delta u &= f && \text{in } \Omega, \\ u &= 0 && \text{at } x = 0 \text{ and } x = 1, \\ \nabla u \cdot \mathbf{n}_{\oplus,\ominus} &= F_{\oplus,\ominus} && \text{at } y = t_{\oplus,\ominus}, \end{aligned}$$

and the right-hand sides of the equation and of the boundary condition are such that

$$u(x, y) = \sin(\pi x) \cdot y^m \quad (m = 1, 2, \dots)$$

is the exact solution.

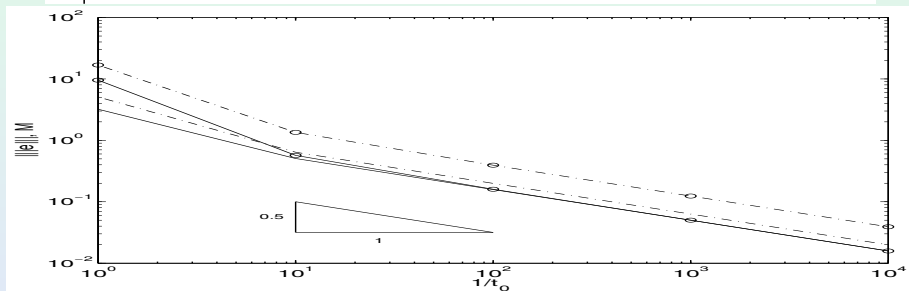
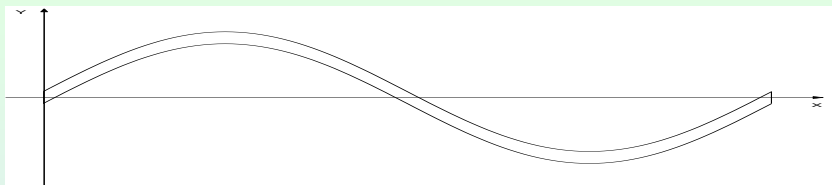


Figure: The domain (top); Convergence rates of the exact error and of the error majorant (bottom); $k = 2$, $m = 4$ (solid lines) and $m = 5$ (dash-dot lines).

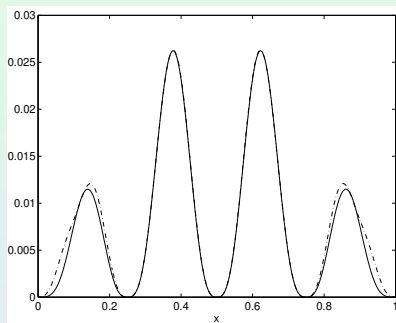
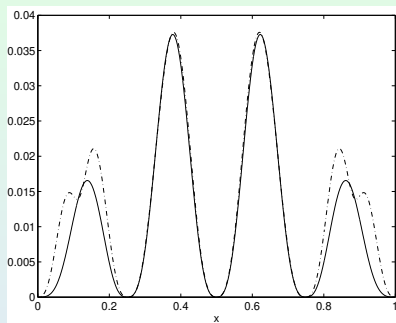


Figure: Distribution of the exact error (solid line) compared with the integrand of the M_1 -term (dash-dot line), $k = 4$, $m = 4$: $t_0 = 0.1$ (top), $t_0 = 0.05$ (bottom).

Penalty type methods

Penalty type methods are often used in numerical analysis and optimal control

There exists a very large amount of publications, e.g., see

- P. Angot, C.-H. Bruneau, P. Fabrie, A penalization method to take into account obstacles in incompressible viscous flows, *Numer. Math.* 81, 497–520, (1999)
- M. Bergounioux A penalization method for optimal control of elliptic problems with state constraints, *SIAM J. Control Optim.*, 30(2), 305–323, (1992).
- R. Glowinski, J.-L. Lions, and R. Trémolierés, *Analyse numérique des inéquations variationnelles*, Dunod, Paris, (1976).
- R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer, New York, 1984.

$$\inf_{v \in K} G(\Lambda v) + (\ell, v)_{\mathcal{V}}$$

Formally, we can encounter the restriction by means of

$$\chi_K(v) := \begin{cases} 0 & \text{if } v \in K, \\ +\infty & \text{if } v \notin K \end{cases}$$

and the variational functional $\mathcal{J}(v) = \mathcal{G}(\Lambda v) + \mathcal{F}(v)$, where

$$\mathcal{F}(v) = \chi_K(v) + (\ell, v)_{\mathcal{V}} \quad \text{and} \quad \ell \in \mathcal{V}.$$

.....
Penalty functional $\Psi : V \rightarrow \mathbb{R}_{\geq 0}$

- $\Psi(v) = 0$ if $v \in K$
- $\Psi(v)$ rapidly grows and tends to $+\infty$ as $\text{dist}(v, K) \rightarrow +\infty$
- $\Psi(v)$ is finite for any $v \in V$

Then for any $\epsilon > 0$ and any $v \in V$ we have

$$\text{regular functional} \quad \boxed{\frac{1}{\epsilon} \Psi(v) \leq \chi_K(v)} \quad \text{jump type functional.} \quad (22)$$

The penalized problem \mathcal{P}_ϵ is to find $u_\epsilon \in V$ such that

$$\mathcal{J}_\epsilon(u_\epsilon) = \inf_{w \in V} \mathcal{J}_\epsilon(w), \quad \mathcal{J}_\epsilon(w) := \mathcal{G}(\Lambda w) + \mathcal{F}_\epsilon(v)$$

$\mathcal{F}_\epsilon = (\ell, v) + \frac{1}{\epsilon} \Psi(v)$ is convex and continuous, so that u_ϵ exists.

.....
It is well known that $u_\epsilon \rightarrow u \subset K$ in V if $\epsilon \rightarrow 0$ provided that the penalty functional Ψ is properly constructed.

Our goal is to get estimates of the distance between u_ϵ and u , which use problem data and u_ϵ (or a numerical approximation of this function).

For this purpose, we again apply the general error estimation theory.

To estimate the error we use the Main Error Identity

$$\mu(v) + \mu^*(y^*) = D_G(\Lambda v, y^*) + D_{\mathcal{F}}(v, -\Lambda^* y^*) \quad (23)$$

We wish to use it for u_ϵ and \mathbf{p}_ϵ^*

$$\mu(u_\epsilon) + \mu^*(\mathbf{p}_\epsilon^*) = D_G(\Lambda u_\epsilon, \mathbf{p}_\epsilon^*) + D_{\mathcal{F}}(u_\epsilon, -\Lambda^* \mathbf{p}_\epsilon^*) \quad (24)$$

$$\mu(u_\epsilon) = D_{\mathcal{F}}(u_\epsilon, -\Lambda^* p^*) + D_G(\Lambda u_\epsilon, p^*), \quad \mu^*(\mathbf{p}_\epsilon^*) = D_{\mathcal{F}}(u, -\Lambda^* \mathbf{p}_\epsilon^*) + D_G(\Lambda u, \mathbf{p}_\epsilon^*).$$

Recall that

$$D_{\mathcal{F}}(v, -\Lambda^* y^*) = \mathcal{F}(v) + \mathcal{F}^*(-\Lambda^* y^*) + \langle \Lambda^* y^*, v \rangle$$

We would like to set here $v = u_\epsilon$, $y^* = p_\epsilon^*$ but $u_\epsilon \notin K$ and we need to use

$$u_\epsilon^K = \pi_K u_\epsilon \quad \pi_K : V \rightarrow K$$

Then (24) yields the general error identity for the errors of penalization:

$$\mu(u_\epsilon^K) + \mu^*(p_\epsilon^*) = \mathcal{D}_G(\Lambda u_\epsilon^K, p_\epsilon^*) + \mathcal{D}_F(u_\epsilon^K, -\Lambda^* p_\epsilon^*).$$

The left hand side is the sum of errors associated with u_ϵ^K and p_ϵ^* :

$$\begin{aligned}\mu(u_\epsilon^K) &:= \mathcal{D}_G(\Lambda u_\epsilon^K, p^*) + \mathcal{D}_F(u_\epsilon^K, -\Lambda^* p^*) \\ \mu^*(p_\epsilon^*) &:= \mathcal{D}_G(\Lambda u, p_\epsilon^*) + \mathcal{D}_F(u, -\Lambda^* p_\epsilon^*).\end{aligned}$$

The right hand side **does not contain u and p^*** and depends only on solutions of the problem \mathcal{P}_ϵ .

.....

Remark: if, G is a quadratic functional, then \mathcal{D}_G are presented by norms, e.g.,

$$\mu(u_\epsilon^K) = \frac{1}{2} \|\Lambda u_\epsilon^K - p^*\|^2 = \frac{1}{2} \|\Lambda(u_\epsilon^K - u)\|^2.$$

We need to compute Fenchel conjugate functions for \mathcal{F} and \mathcal{F}_ϵ . By the definition

$$\mathcal{F}^*(v^*) = \sup_{v \in V} \{(v^* - \ell, v)_V - \chi_K(v)\} = \chi_K^*(v^* - \ell)$$

where χ_K^* is the support functional (convex cone) of the set K

Example:

$$\sup_{\zeta \in (-1,1)} \{(\zeta^* - \ell)\zeta\} = \begin{cases} \zeta^* - \ell & \zeta^* \geq \ell \\ -\zeta^* + \ell & \zeta^* \leq \ell \end{cases}$$

$\mathcal{F}^*(v^*)$ has finite values that are easily computable.

.....

$$\mathcal{F}_\epsilon^*(v^*) = \sup_{v \in V} \{(v^* - \ell, v)_V - \Psi_\epsilon(v)\} = \frac{1}{\epsilon} \Psi^*(\epsilon(v^* - \ell)),$$

Consider the right hand side of the error identity.

Notice that

$$\mathcal{D}_{\mathcal{G}}(\Lambda u_{\epsilon}, p_{\epsilon}^*) = \mathcal{G}(\Lambda u_{\epsilon}) + \mathcal{G}^*(p_{\epsilon}^*) - (p_{\epsilon}^*, \Lambda u_{\epsilon}) = 0$$

We have

$$\begin{aligned} \boxed{\mathcal{D}_{\mathcal{G}}(\Lambda u_{\epsilon}^K, p_{\epsilon}^*)} &= \mathcal{G}(\Lambda u_{\epsilon}^K) + \mathcal{G}^*(p_{\epsilon}^*) - (p_{\epsilon}^*, \Lambda u_{\epsilon}^K) \\ &= \mathcal{G}(\Lambda u_{\epsilon}^K) - \mathcal{G}(\Lambda u_{\epsilon}) + (p_{\epsilon}^*, \Lambda(u_{\epsilon} - u_{\epsilon}^K)), \end{aligned}$$

This part depends on solutions of the penalized problem only.

.....

Example:

$$\frac{1}{2} \|\nabla u_{\epsilon}^K\|_A^2 - \frac{1}{2} \|\nabla u_{\epsilon}\|_A^2 + (A \nabla u_{\epsilon}, \nabla(u_{\epsilon} - \nabla u_{\epsilon}^K)) = \frac{1}{2} \|\nabla(u_{\epsilon} - \nabla u_{\epsilon}^K)\|_A^2$$

$$\boxed{\mathcal{D}_{\mathcal{F}}(u_{\epsilon}^K, -\Lambda^* p_{\epsilon}^*)} = \mathcal{F}(u_{\epsilon}^K) + \mathcal{F}^*(-\Lambda^* p_{\epsilon}^*) + (\Lambda^* p_{\epsilon}^*, u_{\epsilon}^K)$$

Here

$$\mathcal{F}(u_{\epsilon}^K) = \chi_K(u_{\epsilon}^K) + (\ell, u_{\epsilon}^K)_{\mathcal{V}} = (\ell, u_{\epsilon}^K)_{\mathcal{V}}$$

and

$$\mathcal{F}^*(-\Lambda^* p_{\epsilon}^*) = \chi_K^*(-\ell - \Lambda^* p_{\epsilon}^*)$$

Then we find that

$$\mathcal{D}_{\mathcal{F}}(u_{\epsilon}^K, -\Lambda^* p_{\epsilon}^*) = \chi_K^*(-\ell - \Lambda^* p_{\epsilon}^*) + (\ell + \Lambda^* p_{\epsilon}^*, u_{\epsilon}^K)_{\mathcal{V}}.$$

Hence the error identity reads

$$\begin{aligned}\mu(u_\epsilon^K) + \mu^*(p_\epsilon^*) &= \mathcal{G}(\Lambda u_\epsilon^K) - \mathcal{G}(\Lambda u_\epsilon) + (p_\epsilon^*, \Lambda(u_\epsilon - u_\epsilon^K)) \\ &\quad + \chi_K^*(-\ell - \Lambda^* p_\epsilon^*) + (\ell + \Lambda^* p_\epsilon^*, u_\epsilon^K)_V\end{aligned}$$

Notice that

$$(\Lambda^* p_\epsilon^*, u_\epsilon^K)_V = (p_\epsilon^*, \Lambda u_\epsilon^K)$$

Therefore in the right hand side two terms cancel each other

$$\begin{aligned}\dots &= \mathcal{G}(\Lambda u_\epsilon^K) - \mathcal{G}(\Lambda u_\epsilon) + \chi_K^*(-\ell - \Lambda^* p_\epsilon^*) + (p_\epsilon^*, \Lambda u_\epsilon) + (\ell, u_\epsilon^K)_V \\ &= \underbrace{\mathcal{G}(\Lambda u_\epsilon^K) + (\ell, u_\epsilon^K)_V}_{\mathcal{J}(u_\epsilon^K)} - \underbrace{\mathcal{G}(\Lambda u_\epsilon) - (\ell, u_\epsilon)_V}_{\mathcal{J}(u_\epsilon)} \\ &\quad + \chi_K^*(-\ell - \Lambda^* p_\epsilon^*) + (p_\epsilon^*, \Lambda u_\epsilon) + (\ell, u_\epsilon)_V \\ &= \mathcal{J}(u_\epsilon^K) - \mathcal{J}(u_\epsilon) + \chi_K^*(-\ell - \Lambda^* p_\epsilon^*) \\ &\quad + (p_\epsilon^*, \Lambda u_\epsilon) + (\ell, u_\epsilon)_V\end{aligned}$$

We use

$$(p_\epsilon^*, \Lambda u_\epsilon) = (\Lambda^* p_\epsilon^*, u_\epsilon)_V$$

Then

$$\mu(u_\epsilon^K) + \mu^*(p_\epsilon^*) = \mathcal{J}(u_\epsilon^K) - \mathcal{J}(u_\epsilon) + \chi_K^*(-\ell - \Lambda^* p_\epsilon^*) \\ + (\Lambda^* p_\epsilon^* + \ell, u_\epsilon)_V$$

Define

$$\mathcal{R}(p_\epsilon^*) := \ell + \Lambda^* p_\epsilon^*$$

we arrive at the estimate

$$\mu(u_\epsilon^K) + \mu^*(p_\epsilon^*) \leq \mathcal{J}(u_\epsilon^K) - \mathcal{J}(u_\epsilon) + \mathcal{E}(u_\epsilon, p_\epsilon^*)$$

where

$$\mathcal{E}(u_\epsilon, p_\epsilon^*) := \chi_K^*(-\mathcal{R}(p_\epsilon^*)) + \langle \mathcal{R}(p_\epsilon^*), u_\epsilon \rangle$$

Comments:

If the restriction imposed by the set K is inactive (i.e., $u \in \text{int}K$), then $u_\epsilon^K = u_\epsilon = u$ and $p_\epsilon^* = p^*$. In this case, $\Lambda^* p_\epsilon^* + \ell = 0$ and we see that the right hand side vanishes.

.....

How to select optimal projection operator π_K ?

This estimate shows that an "optimal" mapping π_K should generate an element $u_\epsilon^K \in K$ which minimally changes the value of \mathcal{J}_ϵ (in general, the orthogonal projector to K may not satisfy this condition).

Particular case:

If $\mathcal{G}(\Lambda w) = \frac{1}{2}(\mathcal{A}\Lambda w, \Lambda w)$, then the corresponding measures in the right hand side of in terms of norms

$$\mathcal{D}_{\mathcal{G}}(\Lambda u_{\epsilon}^K, p^*) = \frac{1}{2} \|\Lambda(u_{\epsilon}^K - u)\|_{\mathcal{A}}^2$$

$$\mathcal{D}_{\mathcal{G}}(\Lambda u, p_{\epsilon}^*) = \frac{1}{2} \|p^* - p_{\epsilon}^*\|_{\mathcal{A}^{-1}}^2$$

and we arrive at the estimate

$$\frac{1}{2} \|\Lambda(u - u_{\epsilon}^K)\|_{\mathcal{A}}^2 + \frac{1}{2} \|p^* - p_{\epsilon}^*\|_{\mathcal{A}^{-1}}^2 \leq \mathcal{J}(u_{\epsilon}^K) - \mathcal{J}(u_{\epsilon}) + \mathcal{E}(u_{\epsilon}, p_{\epsilon}^*).$$

Validation of mathematical models by comparison with experimental data

Now we discuss the situation typical for engineering and natural sciences.
How to validate a mathematical model using experimental results?

Standard way: we make numerical experiments and compare them with the data. If the results are close, then the model is considered as a suitable one. If not, the model is rejected.

Drawbacks: numerical solution may contain various errors:
approximation,
roundoff,
integration errors,
slow convergence,
instability,
locking,
and **Bugs** in codes.

They may compensate errors of a model, or in opposite they may a good model be looking bad.

If we have a computable measure of a distance to the exact solution of the mathematical model being tested, then a **suitable reconstruction of experimental data can be viewed as an approximation** and directly compared with the solution.

This method does not require explicitly finding the corresponding exact solutions. The data confirm the validity of the model if in all experiments the errors are smaller than the desired tolerance level.

Let u_{\circledast} and p_{\circledast}^* be the functions constructed by experimental measurements. We have:

$$\mu(u_{\circledast}) + \mu(p_{\circledast}^*) = \mathcal{D}_{\mathcal{G}}(\Lambda u_{\circledast}, p_{\circledast}^*) + \mathcal{D}_{\mathcal{F}}(u_{\circledast}, -\Lambda^* p_{\circledast}^*).$$

The right hand side is **computed directly!**

The left hand side shows how far the experimental data are from the theoretical ones.

Example. Suppose the reaction–diffusion model $\boxed{\operatorname{div} A \nabla u - \rho u + f = 0}$ is examined whether it is consistent with the existing set of experimental data: concentration values $u_{*}^{(i)}$ and fluxes $p_{*}^{*(i)}$ obtained in $i = 1, 2, \dots, N_{*}$ experiments with different source terms and other data (e.g., domains). Let $u^{(i)}$ (where $u^{(i)} = u_0^{(i)}$ on Γ) and $p^{*(i)}$ denote the exact solutions of the corresponding problems, **which we do not know and do not try to approximate!** For each experiment, we use the error identity

$$\begin{aligned} & \|\nabla(u^{(i)} - u_{*}^{(i)})\|_A^2 + \rho \|(u^{(i)} - u_{*}^{(i)})\|_{\Omega}^2 + \\ & \|p^{*(i)} - p_{*}^{*(i)}\|_{A^{-1}}^2 + \frac{1}{\rho} \|\operatorname{div}(p^{*(i)} - p_{*}^{*(i)})\|^2 \\ & = \|A \nabla u_{*}^{(i)} - p_{*}^{*(i)}\|_{A^{-1}}^2 + \frac{1}{\rho} \|\operatorname{div} p_{*}^{*(i)} - \rho u_{*}^{(i)} + f^{(i)}\|_{\Omega}^2. \end{aligned}$$

The left hand side is the **error of our model** in the experiment i . The right hand side contains only known experimental data $u_{*}^{(i)}$ and $p_{*}^{*(i)}$. Summarising the results of all experiments we can find an averaged modeling error \tilde{e}_{mod} . It depends of A and ρ . Finding those minimizing \tilde{e}_{mod} , we select the most adequate mathematical model (within the selected class of diffusion type models).

Similar methods based on *estimates of deviations from exact solutions* can be used to estimate **modeling errors** arising in
Deep Neural Networks

Solving PDEs by Deep Neural Networks



W. E, J. Han, A. Jentzen. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations Commun. Math. Stat. (2017) 5:349–380.



Weinan E and Bing Yu. The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems ArXiv:1710.00211v1, 2017.



I. E. Lagaris, A. Likas and D. I. Fotiadis, Artificial neural networks for solving ordinary and partial differential equations, IEEE Transactions on Neural Networks, vol. 9, no. 5, pp. 987-1000.



W. E, B. Yu, The Deep Ritz Method: A deep learning-based numerical algorithm for solving variational problems, Commun. Math. Stat. (2018) 6:1–12.



O. Pironneau, Parameter identification of a fluid-structure system by deep-learning with an Eulerian formulation. Methods Appl. Anal. 26 (2019), no. 3, 281–290.



F. Regazzonia, L.Dede, A.Quarneroni. Machine learning for fast and reliable solution of time-dependent differential equations. Journal of Computational Physics, 397(2019), 108852.



E. Samaniego et al, An energy approach to the solution of partial differential equations in computational mechanics via machine learning. Comput. Methods Appl. Mech. Engrg. 362 (2020), 112790



J. Sirignano and K. Spiliopoulos. DGM: A deep learning algorithm for solving partial differential equations. Journal of Computational Physics, **375**, 1339 – 1364 (2018)

Why do we need to develop AI based methods PDEs ?

Typical answer:

We need them if the standard numerical methods do not work:

- Problems in spaces of high dimension (e.g, economics)
- Stochastic models
- Problems where models and data are rather "flexible" and not fully defined (e.g., biology)

We add another motivation:

PDE based models create excellent testing ground for analysis and development of AI technologies

$$\mathcal{A}u = f, \quad \mathcal{A}: V \rightarrow V',$$

\mathcal{A} is a differential operator and f is a given function (+ other parameters).

Neural Network \mathcal{N} is a surrogate model of the inverse operator \mathcal{A}^{-1} .

$$f \Rightarrow \text{Neural Network} \Rightarrow u_{\mathcal{N}}(x)$$

\mathcal{N} is a graph supplied with weights θ and the so-called "activation function" (usually nonlinear).

It is generated by an iterative "supervised learning" process

$$\mathcal{N}_0 \rightarrow \mathcal{N}_1 \dots \rightarrow \mathcal{N}_k$$





which is essentially based on the so-called

"loss function" $J = J(\mathcal{N})$ and series of known pairs $(f, u(f))$

In terms of the optimal control theory "learning" is a process of *parameter optimisation* (in more complicated cases– *structural optimisation*) using certain "goal functional" J .

Two closely related questions:

Reliability of $u_{\mathcal{N}}$ and efficiency of $J(\mathcal{N})$ for the Learning

-  W. E, B. Yu, The Deep Ritz Method: A deep learning-based numerical algorithm for solving variational problems, Commun. Math. Stat. (2018) 6:1–12.
-  I. E. Lagaris, A. Likas and D. I. Fotiadis, Artificial neural networks for solving ordinary and partial differential equations, IEEE Transactions on Neural Networks, vol. 9, no. 5, pp. 987-1000, 1997.
-  Weinan E and Bing Yu. The Deep Ritz method: A deep learning-based numerical algorithm for solving variational problems, ArXiv:1710.00211v1, 2017.
-  H. Guoa, T. Rabczukb, and X. Zhuang. A Deep Collocation Method for the Bending Analysis of Kirchhoff Plate, arXiv:2102.02617v1.

In the Deep Galerkin the "loss function" $J_{DG}(v)$ is defined as the sum of the residuals of the equation calculated at some set of points.

J_{DG} for the Poisson equation

For the problem

$$\Delta u + f = 0 \text{ in } \Omega, \quad u = u_0 \text{ on } \partial\Omega, \quad (25)$$

we have the loss functional

$$J_{DG}(v) = \frac{1}{N} \sum_{i=1}^N (\Delta v + f)^2|_{x_i}$$

where N – is a number of randomly selected points in Ω .

If $u_0 = 0$, $d = 2$, then $\boxed{\mathcal{N}}$ serves as a **surrogate model** of the well known Green's formula

$$u(x) = \int_{\Omega} G(x, \zeta) f(\zeta) d\zeta \quad G = \frac{1}{2\pi} \ln \rho. \quad (26)$$

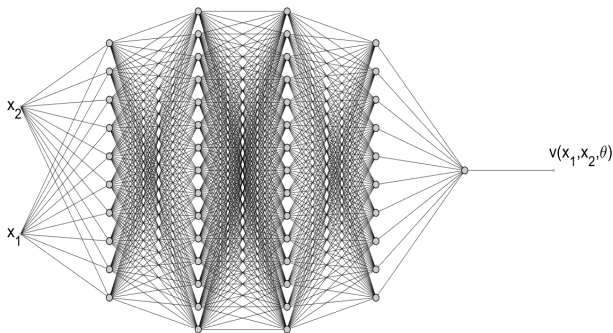
Boundary conditions:

$$v = \alpha(x) \text{Net}(x, \theta) + \beta(x), \quad \alpha|_{\partial\Omega} = 0, \quad \beta|_{\partial\Omega} = u_0,$$

α is a smooth bubble-function vanishing on $\partial\Omega$.

$\boxed{\mathcal{N}}$ has two inputs x_1, x_2 and one output $v(x_1, x_2)$.

"Simple" $\boxed{\mathcal{N}}$



How to verify that $\boxed{\mathcal{N}}$ works correctly?

A natural way is to apply well elaborated methods developed in numerical analysis of PDE's. For this, we need to make a suitable functional counterpart of $u_{\mathcal{N}}$.

$\boxed{\mathcal{N}}$ generates $v_i = u_{\mathcal{N}}(x_i)$ at a given set of points $X_n := \{x_i\} \in \Omega$, $i = 1, 2, \dots, n$. We obtain a mesh-function $\mathbf{v}_n = \{v_i\}$.

Define an extension operator $\Pi : \mathbf{v}_n \rightarrow V$. Let it be

(a) **Consistent**: $\Pi \mathbf{v}_n(x_j) = v_j$ for any point $x_j \in X_n$ and

(b) **Continuous**: from $\mathbf{w}_n \rightarrow \mathbf{v}_n$ in the mesh-norm it follows that

$\|\Pi(\mathbf{w}_n - \mathbf{v}_n)\|_V \rightarrow 0$.

Definition

We say that $u_{\mathcal{N}}$ is ϵ -accurate on X_n , if there exists an extension Π such that

$$\|u - \Pi \mathbf{v}_n\|_V \leq \epsilon = \epsilon_{\mathcal{N}}. \quad (27)$$

First results can be found in:

- Muzalevsky A. V., Repin S. I. A posteriori error control of approximate solutions to boundary value problems constructed by neural networks. Zapiski Nauchn. Semin. Steklov. Inst. Math. (PDMI), v. 499, 77–104, 2020.

It is shown, that formal using of the Deep Galerkin method in certain cases may lead to wrong results.

We verified numerical approximations generated by DNN in a series of different examples and discovered a kind of "locking" phenomenon.

Two simplest examples generated by the problem

$$\Delta u + f = 0, \quad \Omega = (0, 1)^2$$

Problem 1.

$$u_0 = 0, \quad f = -2\pi^2 \sin(\pi x_1) \sin(\pi x_2).$$

Here the exact solution is known, it is a smooth function.

Problem 2.

$$u_0 = |x_1 - 0.5| |x_2 - 0.5|, \quad f = 0.$$

These boundary conditions exclude regular solutions.

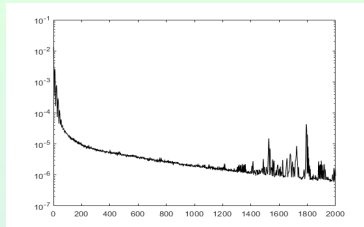
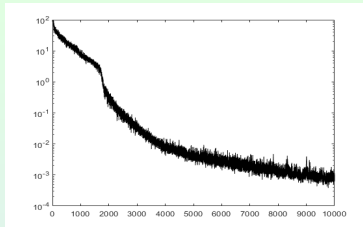


Figure: Network optimization: loss function. Tests 1 (left) and 2 (right)

Test 2: The final loss function is small, i.e. J_{DG} does not indicate any problem.

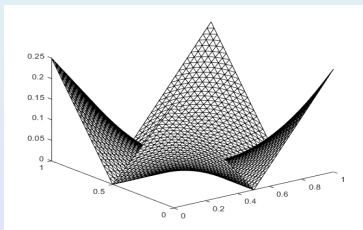
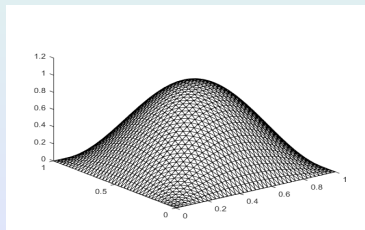


Figure: Tests 1 and 2: DNN solutions

Verification of DNN solutions

h	1/40	1/80	1/160
$\frac{\sqrt{M_{\ominus}(\hat{v})}}{\ \nabla u\ }, \%$	3.92	1.96	0.98
$\frac{\ \nabla(u-\hat{v})\ }{\ \nabla u\ }, \%$	3.92	1.96	0.98
$\frac{\sqrt{M_{\oplus}(\hat{v})}}{\ \nabla u\ }, \%$	7.40	3.70	1.85
$\frac{\sqrt{J_{DG}(v)}}{\ \nabla u\ }, \%$	1.34		
$J_{DG}(v)$	0.00089 OK		

True relative error $\approx 1\%$

h	1/40	1/80	1/160
$\frac{\sqrt{M_{\ominus}(\hat{v})}}{\ \nabla u\ }, \%$	79.05	78.97	78.95
$\frac{\ \nabla(u-\hat{v})\ }{\ \nabla u\ }, \%$	79.07	78.98	78.96
$\frac{\sqrt{M_{\oplus}(\hat{v})}}{\ \nabla u\ }, \%$	90.75	89.96	89.74
$\frac{\sqrt{J_{DG}}}{\ \nabla u\ }, \%$	0.26		
J_{DG}	7.3e-07 OK		

True relative error $\approx 80\%$

Table: Error as function of h . Examples 1 (top) and 2 (bottom).