Записки научных
семинаров ПОМИ
Том 539, 2024 г.

S. Repin

DERIVATION OF FULLY COMPUTABLE ERROR BOUNDS FROM A POSTERIORI ERROR IDENTITIES

Abstract. A posteriori error identities are functional relations that control distances between the exact solution of a boundary value problem and any function from the respective energy space. They have been derived for many boundary value problems associated with partial differential equations of elliptic and parabolic types. A posteriori identities have a common structure: their left hand sides form certain error measures and the right hand ones consist of directly computable terms and a linear functional, which contains unknown error function. Fully computable estimates follow from such an identity provided that this functional is efficiently estimated. The difficulty that arises is due to the fact that computational simplicity and efficiency of such an estimate are contradictory requirements. A method suggested in the paper, largely overcomes this difficulty. It uses an auxiliary finite dimensional problem to estimate the linear functional containing unknown error function. The resulting estimates minimise possible overestimation of this term and imply sharp and fully computable majorants and minorants of errors.

§1. Introduction

As a rule, mathematical models based on partial differential equations operate with so–called weak (generalised) solutions, which are defined as rather abstract objects (elements of infinite-dimensional functional spaces, e.g., Sobolev spaces). If such a problem $\mathcal{A}u = f$ is well–posed, then the solution u exists and is unique, but in the majority of cases there is no hope to find it exactly. Therefore, quantitative analysis actually comes down to the problem of constructing a sufficiently good approximate solution, that is, to the issue of computer modeling. The main limitation is that only discrete (finite-dimensional) objects are representable in a computer.

Key words and phrases: deviations from the exact solution of a boundary value problem, error identities, a posteriori estimates of the functional type.

The research is supported by the Russian Science Foundation (grant 24-21-00293).

Therefore, in quantitative analysis we are forced to replace u by an approximation u_m, which belongs to a certain finite dimensional space V_m ($\dim V_m = m < +\infty$).

Hence the first principal problem is to show that *solutions of finite dimensional problems are able to approximate u with any desired accuracy.* This problem is well studied (e.g., see [4,5,9,12,13,25]). For many classes of problems it is known that u_m tends to u when the dimensionality m tends to infinity provided that certain additional conditions are fulfilled and all the computations are done exactly. These results provide justification of an approximation method. Theoretical convergence guarantees that (in principle) approximations can be found arbitrarily close to the solution.

However, the conditions necessary for convergence and a priori error estimates are often violated. In real life computations, instead of the sequence $\{u_m\}$ we have another sequence $\{\widetilde{u}_m\}$, where $\widetilde{u}_m$ contains errors of various types (e.g., integration and roundoff errors, errors arising in iteration procedures and due to defects in codes). This fact generates the second fundamental problem: *Reliable verification of computational results.* Essentially, it is reduced to finding guaranteed and fully computable estimates of the distance between any function in the admissible functional class (e.g., energy space) and solution of a boundary value problem. The estimate must compute sufficiently sharp bounds of the distance using only robust and well-tested numerical procedures (computation of integrals, solving linear finite dimensional problems, convex minimisation, etc.). If for a class of problems such an estimate is not found, then it remains unclear how to obtain reliable quantitative results. In this situation, we can construct approximations, but have no way to compare them with the exact solution and confirm the validity. Moreover, if it will be shown that for some problem such an estimate is principally impossible, then the respective mathematical model would look like a *thing-in-itself* that admits theoretical considerations but not applicable for quantitative analysis.

In the context of elliptic type problems, the required estimates have beed derived and comprehensively studied over the past 25 years (see [15–18] and references cited therein). All of them follow from the functional relations called *a posteriori error identities.* One side of such an identity contains a measure of the distance between a function(s) computed and the solution. Another side contain integral terms that depend on known data (domain, coefficients, etc.) and known functions. These terms are directly computable. In addition, the identity usually has a different term, which includes unknown error function. The simplest and the most studied example is related to the class of linear boundary value problems of the form (3.1)–(3.2). Here, the corresponding error identity (3.3) has a typical structure. The left hand side of (3.3) is presented by the sum of two squared error norms (where $e := v - u$, $e^* = y - p$, and $p := A\nabla u$). The first term in the right hand side contains known functions v and y (e.g., numerical approximations of u and p), while the second term $(\mathbb{R}(y), e)_{\mathcal{V}}$ contains unknown error function e.

Identities (3.13), (3.26), (3.31), and (3.18) are derived in Sec. 3 for several other problems. They have quite similar structures: error measures in the left hand side and computable quantities in the right hand one. Full computability is violated by the only one term, which is a linear functional of e. Hence, the problem of fully guaranteed and computable error control is de facto reduced to *getting efficient estimates of* $(\mathbb{R}, e)_{\mathcal{V}}$ *by a computable quantity and a certain suitable norm of* e.

Analysis of this problem is the main purpose of the article. In Sec. 4, we suggest several methods to solve it. The corresponding estimates are derived in Sec. 5 and Sec. 6 contains some numerical results that show efficiency of the estimates.

§2. Notation and definitions

First, we recall several notions of convex analysis. Let X be a reflexive Banach space and X^* denote the space conjugate to X with the product $< x^*, x > \in \mathbb{R}$ for $x \in X$ and $x^* \in X^*$. A convex lower semicontinuous function (l.s.c.) functional $\Phi : X \to \mathbb{R}$ has a counterpart $\Phi^* : X^* \to \mathbb{R}$ defined by the relation

$$\Phi^*(x^*) := \sup_{x \in X} \{ < x^*, x > - \Phi(x) \},$$

which is called the *Fenchel conjugate* to Φ. The pair of functionals Φ and Φ^* generates the *compound functional*

$$\mathcal{D}_\Phi (x, x^*) := \Phi(x) + \Phi^*(x^*) - < x^*, x > . \tag{2.1}$$

It is easy to see that

$$\mathcal{D}_\Phi (x, x^*) \geqslant 0 \qquad \forall\, x \in X,\, x^* \in X^*. \tag{2.2}$$

This functional vanishes if and only if x and x^* are connected by certain relations (e.g., see [11])

$$\mathcal{D}_\Phi(x, x^*) = 0 \;\Leftrightarrow\; x^* \subset \partial\Phi(x) \text{ and } x \subset \partial\Phi^*(x^*). \tag{2.3}$$

Here $\partial\Phi$ is the subdifferential of Φ. If Φ (resp. Φ^*) is Gateaux differentiable, then the subdifferential inclusion in (2.3) is replaced by the relation $x^* = \Phi'(x)$ (resp. $x = \Phi^{*\prime}(x^*)$), where prime denotes the derivative.

Throughout the paper V denotes a reflexive Banach space with the norm $\|\cdot\|_V$ and V^* is the space conjugate to V. The duality pairing of $v \in V$ and $v^* \in V^*$ is denoted by $\langle v^*, v\rangle$. $\mathcal{V}$ is a Hilbert space with the scalar product $(\cdot,\cdot)_{\mathcal{V}}$. It is assumed that V is is compactly embedded in $\mathcal{V}$, so that the spaces V, $\mathcal{V}$, and V^* form the so called Helfand triple. If $v^* \in \mathcal{V}$ then $\langle v^*, v\rangle = (v^*, v)_{\mathcal{V}}$.

We use another Hilbert space U supplied with the scalar product $(\cdot,\cdot)_U$ and the norm $\|\cdot\|_U$. Next, $\Lambda : V \to U$ is a bounded linear operator and $\Lambda^* : U \to V^*$ is the conjugate operator defined by the relation

$$(y, \Lambda v)_U = \langle \Lambda^* y, v\rangle \qquad \forall y \in U,\ v \in V.$$

Also, we introduce the subspace

$$Q^* := \{q \in U \mid \Lambda^* q \in \mathcal{V}\}.$$

Let $A : U \to U$ be a bounded positive definite operator and A^{-1} denote the respective inverse operator. Using them we introduce the spaces Y and Y^* that contain the same elements as U but operate with different norms $\|y\|_A^2 := (Ay, y)_U$ and $\|y\|_{A^{-1}}^2 := (A^{-1}y, y)_U$. It is assumed that there exist constants $0 < \underline{c}_A \leqslant \overline{c}_A$ such that

$$\underline{c}_A \|y\|_U^2 \leqslant \|y\|_A^2 \leqslant \overline{c}_A \|y\|_U^2 \qquad \forall y \in U. \tag{2.4}$$

Elements of spaces V, $\mathcal{V}$, and U are functions defined in an open bounded domain $\Omega \subset \mathbb{R}^d$, $d \geqslant 1$ with Lipschitz boundary Γ. In the examples below, they are Lebesgue and Sobolev spaces for scalar and vector-valued functions. For them, we use standard notation $L_p(\Omega)$ (or $L_p(\Omega, \mathbb{R}^d)$) and $W_p^l(\Omega)$ (where $l, p \geqslant 1$) and mark above by $\circ$ if the respective functions vanish on Γ. Norms of scalar and vector valued functions, which are squre integrable in Ω are denoted by $\|\cdot\|_\Omega$.

If KerΛ contains only zero function, then $\|\Lambda w|_U$ can be used as a norm of V and we have the inequality

$$\|w\|_{\mathcal{V}} \leqslant C_\Lambda \|\Lambda w\|_U \qquad \forall w \in V. \tag{2.5}$$

For example, if Λ is the gradient operator and V contains functions vanishing near the boundary, then (2.5) is the Friedrichs inequality. The constant C_Λ reflects important quantitative relations associated with the spaces $\mathcal{V}$ and V. Therefore, it often arises in a posteriori estimates of the functional type. However, we can also use other inequalities that can be viewed as advanced forms of (2.5). Such an inequality has the form

$$\inf_{\psi\in\Psi}\|w-\psi\|_{\mathcal{V}}\leqslant C_{\Psi,\Lambda}\|\Lambda w\|_U \qquad \forall w\in V_\Psi, \tag{2.6}$$

where $\Psi\in\mathcal{V}$ is a certain set of functions that excludes any function from KerΛ. In this case, the constant $C_{\Psi,\Lambda}$ is usually smaller than C_Λ, so that it may be advantageous to use (2.6) instead of (2.5). The simplest form of (2.6) is known as the Poincaré inequality for functions with zero mean:

$$\inf_{c\in\mathbb{R}}\|w-c\|_\Omega\leqslant C_P\|\nabla w\|_\Omega \qquad \forall w\in \overset{\circ}{H}{}^1(\Omega). \tag{2.7}$$

If Ω is a convex domain, then $C_P\leqslant\frac{1}{\pi}\mathrm{diam}\,\Omega$ (see [14]).

§3. A posteriori error identities

A posteriori error identities are functional relations that contain measures of errors in one side and computable quantities in the other. They hold for any functions that belong to the basic (energy) class. Therefore, error identities form a basis for getting fully computable error estimates for a wide spectrum of approximations regardless of their origin.

3.1. Linear problems of the type $\Lambda^*A\Lambda u+\ell=0$. This class of mathematical models originates from the equations

$$\Lambda^*p+\ell=0, \tag{3.1}$$

$$p=A\Lambda u, \tag{3.2}$$

which usually reflect physical relations: (3.1) is a certain conservation (balance) law and (3.2) is a constitutive relation (physical law associated with a particular media). Hence the solution is presented by two functions u and p. Let $v\in V$ be a function considered as an approximation of u and $y\in U$ be an approximation of p. Then

$$e:=v-u \quad\text{and}\quad e^*:=y-p$$

are the corresponding errors.

Error identities for problems of the class (3.1)–(3.2) are well studied (see [18–20]). If $\ell\in\mathcal{V}$, then the identity reads as follows:

$$\|e^*\|^2_{A^{-1}}+\|\Lambda e\|^2_A=\|y-A\Lambda v\|^2_{A^{-1}}+2(\mathbb{R}(y),e)_{\mathcal{V}}, \tag{3.3}$$

where

$$\mathbb{R}(y) := \Lambda^* y + \ell \in \mathcal{V}.$$

Applications of (3.3) to a posteriori error estimation of numerical approximations was studied in [20]. It can be also used to analysis of modeling errors (see [19, 23] and some other publications cited therein).

Consider a more general case, where $\ell \in V^*$ and has the form

$$\langle \ell, w \rangle = (z, \Lambda w)_U + (f, w)_{\mathcal{V}}, \quad \text{where } f \in \mathcal{V} \text{ and } \quad z \in U.$$

Then the corresponding generalised solution is defined by the relation

$$(A\Lambda u + z, \Lambda w)_U + (f, w)_{\mathcal{V}} = 0 \qquad \forall w \in V. \tag{3.4}$$

In this case, $p := A\Lambda u + z$ and (3.4) reads

$$(p, \Lambda w)_U + (f, w)_{\mathcal{V}} = 0 \qquad \forall w \in V.$$

Hence for any $v \in V$ and $y \in Q^*$,

$$\begin{aligned}\|y - A\Lambda v - z\|^2_{A^{-1}} &= \|y - p\|^2_{A^{-1}} + \|\Lambda(v - u)\|^2_A - 2(y - p, \Lambda(v - u))_U \\ &= \|y - p\|^2_{A^{-1}} + \|\Lambda(v - u)\|^2_A - 2(\Lambda^* y + f, v - u)_{\mathcal{V}}\end{aligned}$$

and we arrive at the identity

$$\|e^*\|^2_{A^{-1}} + \|\Lambda e\|^2_A = \|y - A\Lambda v - z\|^2_{A^{-1}} + 2(\mathbb{R}(y), e)_{\mathcal{V}}, \tag{3.5}$$

where

$$\mathbb{R}(y) := \Lambda^* y + f \in \mathcal{V}.$$

We can rewrite (3.5) in the following form:

$$\begin{aligned}\|\Lambda e\|^2_A + \|e^*\|^2_{A^{-1}} &= \|y - A\Lambda v + z\|^2_{A^{-1}} \\ &\quad + 2\gamma(e^*, \Lambda e) + 2(1 - \gamma)(\mathbb{R}(y), e)_{\mathcal{V}}, \quad \gamma \in [0, 1]. \end{aligned}\tag{3.6}$$

By the Young's inequality

$$2\gamma|(e^*, \Lambda e)| \leqslant \gamma\epsilon\|\Lambda e\|^2_A + \frac{\gamma}{\epsilon}\|e^*\|^2_{A^{-1}}, \quad \epsilon > 0,$$

we obtain the estimate

$$\begin{aligned}(1 - \gamma\epsilon)\|\Lambda e\|^2_A + \left(1 - \frac{\gamma}{\epsilon}\right)\|e^*\|^2_{A^{-1}} \\ \leqslant \|y - A\Lambda v - z\|^2_{A^{-1}} + 2(1 - \gamma)(\mathbb{R}(y), e)_{\mathcal{V}}, \end{aligned}\tag{3.7}$$

where $\gamma \leqslant \epsilon \leqslant \frac{1}{\gamma}$.

Two limit versions of (3.7) (for $\epsilon = \gamma$ and $\epsilon = \frac{1}{\gamma}$) imply the estimates

$$(1-\gamma^2)\max\Big\{\|\Lambda e\|^2_A, \|e^*\|^2_{A^{-1}}\Big\} \leqslant \|y - A\Lambda v - z\|^2_{A^{-1}} + 2(1-\gamma)(\mathbb{R}(y), e)_{\mathcal{V}} \quad (3.8)$$

and

$$(1+\gamma\epsilon)\|\Lambda e\|^2_A + \big(1+\tfrac{\gamma}{\epsilon}\big)\, \|e^*\|^2_{A^{-1}} \geqslant \|y - A\Lambda v - z\|^2_{A^{-1}} + 2(1-\gamma)(\mathbb{R}(y), e)_{\mathcal{V}}, \quad (3.9)$$

where $\epsilon > 0$. In particular, for $\gamma = \epsilon = 1$ we have

$$\|\Lambda e\|^2_A + \|e^*\|^2_{A^{-1}} \geqslant \frac{1}{2}\|y - A\Lambda v - z\|^2_{A^{-1}}. \quad (3.10)$$

The first term in the right hand sides of (3.5), (3.7), and (3.9) depends on known data (A, z, Ω) and computed functions v and y. It is directly computable. However, the second term $(\mathbb{R}(y), e)_{\mathcal{V}}$ contains unknown error function e. This situation is typical for many other problems.

3.2. Convection diffusion problem. As a second example, we consider the convection diffusion problem

$$-\Delta u + \mathrm{a}\cdot\nabla u = f \qquad \text{in } \Omega, \quad (3.11)$$

$$u = 0 \qquad \text{on } \Gamma. \quad (3.12)$$

Let

$$\mathrm{a} \in L_\infty(\Omega, \mathbb{R}^d) \ \text{ and } \ \mathrm{div}\,\mathrm{a} = 0.$$

In this case, $U = L_2(\Omega, \mathbb{R}^d)$, $\mathcal{V} = L_2(\Omega)$, $V = \overset{\circ}{H}{}^1(\Omega)$, and $p = \nabla u$. It is easy to see that

$$\begin{aligned}\|\nabla v - y\|^2_\Omega &= \|\nabla e\|^2_\Omega + \|e^*\|^2_\Omega + 2\int_\Omega \nabla e\cdot e^* dx \\ &= \|\nabla e\|^2_\Omega + \|e^*\|^2_\Omega - 2\int_\Omega e\,\mathrm{div} e^* dx \\ &= \|\nabla e\|^2_\Omega + \|e^*\|^2_\Omega - 2\int_\Omega e(\mathrm{div} y - \mathrm{a}\cdot p + f) dx.\end{aligned}$$

Since

$$\int_\Omega e\,\mathrm{a}\cdot\nabla e dx = \frac{1}{2}\int_\Omega \mathrm{a}\cdot\nabla e^2\,dx = 0,$$

we conclude that

$$\|\nabla e\|^2_\Omega + \|e^*\|^2_\Omega = \|\nabla v - y\|^2_\Omega - 2\int_\Omega \mathbb{R}(y,v)\,e\,dx, \tag{3.13}$$

where

$$\mathbb{R}(y,v) := \mathrm{div} y - \mathrm{a}\cdot\nabla v + f.$$

This error identity has the same structure as (3.3). The left hand side is a natural error measure and the right hand one consists of the computable norm $\|\nabla v - y\|$ and the functional $\int_\Omega \mathbb{R}(y,v)\,edx$. It is clear that tis integral term can be estimated in such a way that the unknown e is estimated by the norm $\|\nabla e\|$, so that the identity implies fully computable error estimates. The problem is how to do this efficiently avoiding significant overestimation. Various options are discussed in Sec. 4.

Identities similar to (3.13) hold for the evolutionary convection-diffusion problem

$$u_t - \mathrm{div} p + \mathrm{a}\cdot\nabla u + \varrho^2 u - f \qquad \text{in } Q_T := \Omega\times(0,T), \tag{3.14}$$

$$u(x,t) = 0 \qquad \text{in } S_T := \Gamma\times(0,T), \tag{3.15}$$

$$u(x,0) = u_0(x) \qquad x\in\Omega, \tag{3.16}$$

$$p = A\nabla u \qquad \text{in } Q_T := \Omega\times(0,T) \tag{3.17}$$

with positive definite symmetric matrix A and coefficients satisfying the conditions

$$\mathrm{a} = L_\infty(\Omega,\mathbb{R}^d),\ \mathrm{div a}\in L_\infty(\Omega),\ 0\leqslant -\frac{1}{2}\mathrm{div a} + \varrho^2 =: \sigma_a^2,$$

$$0<\varrho\in L_\infty(\Omega),\quad f\in L_2(Q_T),\quad u_0\in \overset{\circ}{H}{}^1(\Omega).$$

In [22], it is shown that for the problem (3.14)-(3.17) the following identity holds:

$$\boldsymbol{\mu}(e,e^*) + \|e(x,T)\|^2_\Omega = \|e(x,0)\|^2_\Omega + \int_0^T \|y - A\nabla v\|_{A^{-1}}dt - 2\int_{Q_T}\mathbb{R}_f(v,y)edxdt, \tag{3.18}$$

where $v(x,t)$ and $y(x,t)$ are approximations of $u(x,t)$ and $p(x,t)$, respectively,

$$\boldsymbol{\mu}(e,e^*) := \left(\int\limits_0^T (\|\nabla e\|_A^2 + \|e^*\|_{A^{-1}}^2 + 2\|\sigma_a e\|_\Omega^2)dt\right)^{1/2}$$

is a measure of deviation from (u,p), and

$$\mathbb{R}_f(v,y) := f - v_t + \operatorname{div} y - \mathrm{a}\cdot\nabla v - \varrho^2 v$$

can be viewed as residual of (3.14). If $v(x,t)$ satisfies the initial condition (3.16), then the first term in the right hand side of (3.18) vanishes. The second term contains known functions v and y and can be easily computed. The third term is the integral formed by unknown e and known residual function $\mathbb{R}_f(v,y)$. We see that the structure of (3.18) is the same as of (3.3) and (3.13).

3.3. General elliptic problem for monotone operators. Consider the following abstract elliptic problem: find $u \in V$, $p^* \in Y^*$ and $\sigma \in V^*$, such that

$$\mathcal{A}u + \ell = 0, \qquad \text{where} \quad \mathcal{A}u := \Lambda^* p^*(u) + \omega\sigma(u), \tag{3.19}$$

where $\ell \in V^*$, $\omega \geqslant 0$. The dependence of p^* from u is defined implicitly (cf. (2.1)–(2.3)) by the relation

$$\mathcal{D}_G(\Lambda u, p^*) := G(\Lambda u) + G^*(p^*) - (p^*, \Lambda u) = 0. \tag{3.20}$$

If G is differentiable, then (3.20) implies the explicit relation

$$p^* = G'(u).$$

By means of (3.20) we may consider a various functional relations (including those presented by multivalued mappings) in a compact unified form.

Analogously, $\sigma = \sigma(u)$ is defined by the relation

$$\mathcal{D}_R(u,\sigma) := \quad R(u) + R^*(\sigma) - \langle\sigma, u\rangle = 0. \tag{3.21}$$

We assume that the functionals $G : Y \to \mathbb{R}$ and $R : V \to \mathbb{R}$ are convex, continuous, and nonnegative functionals. In addition

$$G(0_Y) = R(0_V) = 0, \tag{3.22}$$

where 0_Y and 0_V are zero elements of Y and V, respectively. Also, we assume that the functional G satisfies the coercivity condition

$$\liminf \frac{G(\Lambda w)}{\|w\|_V} = +\infty \qquad \text{as } \|w\|_V \to +\infty. \tag{3.23}$$

Notice that in this general case, the spaces Y and Y^* contain different elements and $(\cdot, \cdot)$ denotes the duality pairing of these spaces. Therefore, in this section we mark elements of Y^* by stars.

Lemma 1. *The operator $\mathcal{A}$ defined by* (3.20), (3.21)*, and* (3.23) *is monotone and coercive.*

Proof. Let $u_1, u_2 \in V$, $\sigma_1, \sigma_2 \in V^*$, and $p_1^*, p_2^* \in Y^*$ satisfy the conditions

$$\mathcal{D}_G(\nabla u_i, p_i^*) = 0 \quad \text{and} \quad \mathcal{D}_R(u_i, \sigma_i) = 0 \quad i = 1, 2. \tag{3.24}$$

We have

$$\langle \mathcal{A}u_1 - \mathcal{A}u_2, u_1 - u_2 \rangle = (p_1^* - p_2^*, \Lambda(u_1 - u_2)) + \langle \sigma_1 - \sigma_2, u_1 - u_2 \rangle.$$

In view of (3.20) and (3.21), the right hand side of this identity is equal to

$$G(\nabla u_1) + G^*(p_1^*) + G(\nabla u_2) + G^*(p_2^*) - (p_2^*, \nabla u_1) - (p_1^*, \nabla u_2)$$
$$+ R(u_1) + R^*(\sigma_1) + R(u_2) + R^*(\sigma_2) - \langle \sigma_2, u_1 \rangle - \langle \sigma_1, u_2 \rangle.$$

Hence

$$\langle \mathcal{A}u_1 - \mathcal{A}u_2, u_1 - u_2 \rangle$$
$$= \mathcal{D}_G(\nabla u_1, p_2^*) + \mathcal{D}_G(\nabla u_2, p_1^*) + \mathcal{D}_R(u_1, \sigma_2) + \mathcal{D}_R(u_2, \sigma_1).$$

Recalling (2.2), we see that the operator $\mathcal{A}$ is monotone.

Since

$$\langle \mathcal{A}w, w \rangle = \langle \Lambda^* p^*(w) + \omega\sigma(w), w \rangle = (p^*(w), \Lambda w)_U + \omega \langle \sigma(w), w \rangle$$

and

$$G(\Lambda w) + G^*(p^*(w)) - (p^*(w), \Lambda w)_U = 0,$$
$$R(w) + R^*(\sigma(w)) - \langle \sigma(w), w \rangle = 0,$$

we see that

$$\langle \mathcal{A}w, w \rangle = G(\Lambda w) + \omega R(w) + G^*(p^*(w)) + \omega R^*(\sigma(w)).$$

In view of (3.22), congugate functionals satisfy the condition

$$G^*(p^*) + \omega R^*(\sigma) \geqslant 0.$$

Hence by (3.23) we find that

$$\liminf \frac{\langle \mathcal{A}w, w\rangle}{\|w\|_V} = +\infty \qquad \text{as } \|w\|_V \to +\infty$$

and the coercivity is guaranteed. □

Well-posedness of the problem (3.19)–(3.21) follows from Lemma 1 and Browder-Minty theorem. The corresponding generalised solution satisfies the relation

$$(p^*(u), \Lambda w) + \omega\langle \sigma(u), w\rangle + \langle \ell, w\rangle = 0 \quad \forall w \in V, \tag{3.25}$$

where $p^*(u)$ and $\sigma(u)$ are defined by (3.20) and (3.21).

A posteriori error identity for this class of problems was derived in [21]. For convenience of the reader, we reproduce (with some modifications) the respective proof below.

Theorem 1. *For any $v \in V$, $y^* \in Y^*$, and $\tau \in V^*$, it holds*

$$\mathcal{D}_G(\Lambda u, y^*) + \mathcal{D}_G(\Lambda v, p^*) + \omega\mathcal{D}_R(u, \tau) + \omega\mathcal{D}_R(v, \sigma) = \mathcal{D}_G(\Lambda v, y^*) + \omega\mathcal{D}_R(v, \tau) + \langle \mathbb{R}(y^*, \tau), e\rangle, \tag{3.26}$$

where $\mathbb{R}(y^*, \tau) := \Lambda^* y^* + \omega\tau + \ell$.

Proof. It is easy to see that

$$\begin{aligned}\mathcal{D}_G(\Lambda v, y^*) &= G(\Lambda u) + G^*(y^*) - (y^*, \Lambda u) + G(\Lambda v) + G^*(p^*) - (p^*, \Lambda v) \\ &\quad + (y^*, \Lambda u) + (p^*, \Lambda v) - (p^*, \Lambda u) - (y^*, \Lambda v) \\ &= \mathcal{D}_G(\Lambda u, y^*) + \mathcal{D}_G(\Lambda v, p^*) + (p^* - y^*, \Lambda(v - u)) \end{aligned} \tag{3.27}$$

and

$$\begin{aligned}&\mathcal{D}_R(v, \tau) = R(v) + R^*(\tau) - \langle \tau, v\rangle \\ &= R(u) + R^*(\tau) - \langle \tau, u\rangle + R(v) + R^*(\sigma) - \langle \sigma, v\rangle + \langle \tau, u\rangle + \langle \sigma, v\rangle - \langle \sigma, u\rangle - \langle \tau, v\rangle \\ &= \mathcal{D}_R(u, \tau) + \mathcal{D}_R(v, \sigma) + \langle \sigma - \tau, v - u\rangle. \end{aligned} \tag{3.28}$$

By (3.27) and (3.28) we obtain

$$\begin{aligned}&\mathcal{D}_G(\Lambda v, y^*) + \omega\mathcal{D}_R(v, \tau) \\ &\quad = \mathcal{D}_G(\Lambda u, y^*) + \mathcal{D}_G(\Lambda v, p^*) + \omega\mathcal{D}_R(u, \tau) + \omega\mathcal{D}_R(v, \sigma) \\ &\qquad + (p^* - y^*, \Lambda(v - u)) + \omega\langle \sigma - \tau, v - u\rangle. \end{aligned} \tag{3.29}$$

Since $e \in V$, we use (3.25) and find that

$$(p^*, \Lambda e) + \omega\langle\sigma, e\rangle = -\langle\ell, e\rangle.$$

Using this relation we represent the last two terms of (3.29) in the form

$$(p^* - y^*, \Lambda e) + \omega\langle\sigma - \tau, e\rangle = \langle -\Lambda^* y^* - \omega\tau - \ell, e\rangle = -\langle\mathbb{R}(y^*, \tau), e\rangle. \tag{3.30}$$

Now (3.26) follows from (3.29) and (3.30). □

Remark 1. If $\omega = 0$, then we arrive at the identity

$$\mathcal{D}_G(\Lambda u, y^*) + \mathcal{D}_G(\Lambda v, p^*) = \mathcal{D}_G(\Lambda v, y^*) + \langle\Lambda^* y^* + \ell, e\rangle. \tag{3.31}$$

Applications of this type error identities to estimation of errors of generated by simplification, homogenisation, and dimension reduction of mathematical models are studied in the book [19].

First versions of error identities were derived for convex variational problems, where we can use properties of the primal and dual variational problems [15, 16]. They are derived for various nonlinear problems (see [2, 24] and a systematic overview in the book [19]). However, error identities also hold for problems that have no variational settings (see [21, 22] and some other publications cited therein). In the proof of Theorem 1 we have used (3.25) and properties of compound functionals only. Hence justification of (3.26) (and other above presented identities) do not require variational duality arguments.

We see that the identity (3.26) has the same structure as (3.3), (3.5), (3.13) and many others derived in [2, 16, 19, 21, 24] for various nonlinear problems. The left hand side of (3.26) consist of four nonnegative terms, which can be viewed as nonlinear error measures. They vanish if and only if approximations coincide with the exact solutions, while the right hand one contains directly computable terms $\mathcal{D}_G(\Lambda v, y^*)$ and $\omega\mathcal{D}_R(v, \tau)$ together with the term $\langle\mathbb{R}(y^*, \tau), e\rangle$.

§4. Evaluation of the term $\langle\mathbb{R}, e\rangle$

Identities (3.5), (3.13), (5.11), (3.31), and (3.26) contain the term $\langle\mathbb{R}, e\rangle$, where the function e is unknown. If we knew the value of this term, then the error control problem would be completely solved. Therefore, the key question is how to estimate $\langle\mathbb{R}, e\rangle$ via the quantities in the left hand sides of the identities. In this section, we discuss various ways of getting such type estimates. They develop the ideas earlier exposed in [20].

4.1. Computability and efficiency. First, we note that

$$|\langle\, \mathbb{R}, e\,\rangle| \;\leqslant \|\mathbb{R}\|_{V^*}\|\Lambda e\|_U, \tag{4.1}$$

where

$$\|\mathbb{R}\|_{V^*} = \sup_{w\in V}\frac{\langle \mathbb{R}, w\rangle}{\|\Lambda w\|_U}.$$

If $\mathbb{R}$ is more regular and belongs to $\mathcal{V}$, then using (2.5) we have

$$\|\mathbb{R}\|_{V^*} = \sup_{w\in V}\frac{(\mathbb{R}, w)_{\mathcal{V}}}{\|\Lambda w\|_U} \leqslant C_\Lambda\|\mathbb{R}\|_{\mathcal{V}}$$

and obtain the estimates

$$|\langle\, \mathbb{R}, e\,\rangle| \;\leqslant C_\Lambda\|\mathbb{R}\|_{\mathcal{V}}\|\Lambda e\|_U \tag{4.2}$$

and

$$|\langle\, \mathbb{R}, e\,\rangle| \;\leqslant \frac{C_\Lambda^2}{2\mu \underline{c}_A}\|\mathbb{R}\|_{\mathcal{V}}^2 + \frac{\mu}{2}\|\Lambda e\|_A^2 \qquad \forall \mu > 0. \tag{4.3}$$

Setting $\mu \in (0,1]$ and applying (4.1) and (4.3) to the simplest error identity (3.3), we get estimates for the combined error norms with weights:

$$(1-\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 \leqslant \|y - A\Lambda v\|_{A^{-1}}^2 + \frac{1}{\mu \underline{c}_A}\|\mathbb{R}\|_{V^*}^2 \tag{4.4}$$

and

$$(1-\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 \leqslant \|y - A\Lambda v\|_{A^{-1}}^2 + \frac{C_\Lambda^2}{\mu \underline{c}_A}\|\mathbb{R}\|_{\mathcal{V}}^2, \tag{4.5}$$

The estimate (4.4) is theoretically correct, but practically useless because it operates with an incomputable supremum type norm. The norm $\|\mathbb{R}\|_{\mathcal{V}}$ is computable (it is an intergal type norm), so that the right hand side of (4.4) is easy to calculate. However, (4.2) is a much coarser upper bound than (4.1). Besides, there is another difference between $\|\mathbb{R}\|_{V^*}$ and $\|\mathbb{R}\|_{\mathcal{V}}$ essential from the viewpoint of numerical applications. As a rule, approximations (e.g., Galerkin finite element approximations u_h) converge to the exact solution u in the basic energy space only, i.e., $\|u_h - u\|_V \to 0$ as the mesh parameter h tends to zero. If y is defined by a simple reconstruction of u_h (e.g., $y_h = A\Lambda u_h$), then $\mathbb{R}(y_h)$ may not belong to $\mathcal{V}$ and converges in a weak sense only, i.e.,

$$\|\mathbb{R}(y_h)\|_{V^*} \to 0. \tag{4.6}$$

For (4.4) this type of convergence is admissible, but for the estimate (4.5) it is too weak. Various averaging (post–processing) procedures are often applied to replace y_h by a close and more regular function $\widetilde{y}_h$. The resulting improvements are usually reduced to the following: we have uniform boundedness of $\|\mathbb{R}(\widetilde{y}_h)\|_{\mathcal{V}}$ and can show that

$$\mathbb{R}(\widetilde{y}_h) \to 0 \quad \text{weakly in } \mathcal{V} \text{ as } h \to 0. \tag{4.7}$$

In this case, (4.5) can be used for the pair $(v, \widetilde{y}_h)$, but it may essentially overestimate the error.

In other words, the estimate (4.4) is fully adapted to properties of numerical approximations converging to the solution pair (u, p). However, its right hand side contains an incomputable term. In opposite, (4.5) is fully computable, but may generate essential overestimation and in certain cases may be inefficient.

To overcome this contradiction between the efficiency and computability we follow the idea suggested in [20]. It is based on *using an auxiliary problem generated by* $\mathbb{R}$. Let $u_{\mathrm{R}} \in V$ be such that

$$(\mathrm{B}\Lambda u_{\mathrm{R}}, \Lambda w)_U = \langle\, \mathbb{R}, w\, \rangle, \qquad \forall w \in V, \tag{4.8}$$

where $\mathrm{B} : U \to U$ is a certain bounded self-adjoint positive definite operator satisfying the condition

$$\underline{c}_{\mathrm{B}} \|y\|_U^2 \leqslant (\mathrm{B}y, y)_U =: \|y\|_{\mathrm{B}}^2 \leqslant \overline{c}_{\mathrm{B}} \|y\|_U^2.$$

This problem is uniquely solvable for any $\mathbb{R} \in V^*$ and the operator B is at our disposal. We may set $\mathrm{B} = A$, or define it as a simplification of A, or even set $\mathrm{B} = \mathbb{1}$, where $\mathbb{1}$ is the unit operator. By (4.8) we obtain

$$\langle\, \mathbb{R}, e\, \rangle \leqslant \|\Lambda u_{\mathrm{R}}\|_U \|\Lambda e\|_U$$

and replace (4.4) by

$$(1-\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 \leqslant \|y - A\Lambda v\|_{A^{-1}}^2 + \frac{1}{\mu \underline{c}_A} \|\Lambda u_{\mathrm{R}}\|_U^2. \tag{4.9}$$

Let $\widetilde{y}_h$ be a sequence satisfying (4.6) and $\mathbb{R}(\widetilde{y}_h) \in \mathcal{V}$. Then the right hand side of (4.8) tends to zero and, therefore, $\|\Lambda u_{\mathbb{R}(\widetilde{y}_h)}\|_{\mathcal{V}} \to 0$. Hence proper behaviour of the right hand side of (4.9) is guaranteed. Of course, this fact is not sufficient to say that all difficulties have been overcome. The function u_{R} that solves (4.6) is generally unknown. However, we get a computable estimate if to replace (4.6) by a *finite dimensional problem:*

Find $u_{\mathrm{R},m} \in V_m \subset V$

$$(\mathrm{B}\Lambda u_{\mathrm{R},m}, \Lambda w_i)_U = \langle \mathbb{R}, w_i \rangle \qquad \forall w_i \in V_m, \tag{4.10}$$

where $w_i \in V$, $i = 1, 2, ..., m$ are linearly independent functions and

$$V_m := \mathrm{span}\{w_1, w_2, ..., w_m\}.$$

For any function $v \in V$, we define the orthogonal projection on V_m as the element $v_m \in V_m$ satisfying the relations

$$(\mathrm{B}\Lambda(v_m - v), \Lambda w_i)_U = 0 \qquad \forall\, w_i \in V_m. \tag{4.11}$$

Let $V_m^\perp$ be the orthogonal complement to V, i.e.,

$$V_m^\perp = \{v \in V \mid (\mathrm{B}\Lambda v, \Lambda w_i)_U = 0, \quad i = 1, 2, ..., m\}.$$

In view of (4.11), $e_m^\perp := e - e_m \in V_m^\perp$ and by (4.10) we have the principal decomposition

$$\langle \mathbb{R}, e \rangle = (\mathrm{B}\Lambda u_{\mathrm{R},m}, \Lambda e_m)_U + \langle \mathbb{R}, e_m^\perp \rangle. \tag{4.12}$$

The first term in the right hand side of (4.12) tends to zero if $\mathbb{R}$ weakly tends to zero (this fact follows from (4.10)) and $u_{\mathrm{R},m}$ is known (it is found by solving a finite dimensional problem). Therefore, this term satisfies the requirements. The second term contains not the whole error e, but only its projection on $V_m^\perp$. If m grows then the impact of this term decreases. Our next goal is to estimate it as accurate as possible using properties of specially constructed subspaces V_m.

4.2. Decomposition of $\langle \mathbb{R}, e\rangle$ by a set of orthogonal functions. Without a loss of generality, we may assume that the functions w_i, $i = 1, 2, ..., m$ are orthogonalised and normed, so that

$$(\mathrm{B}\Lambda w_i, \Lambda w_j)_U = \delta_{ij}, \tag{4.13}$$

where δ_{ij} is the Kronecker symbol. Using the Gram–Schmidt orthonormalization a system of m linearly independent functions can be reformed to a system satisfying (4.13). Another condition imposed on w_i is more demanding: we assume that the functions w_i are sufficiently regular, so that

$$g_i := \mathcal{B} w_i \, \in \mathcal{V} \qquad i = 1, 2, ..., m, \tag{4.14}$$

where $\mathcal{B} : V \to V^*$ is defined as $\mathcal{B} = \Lambda^* \mathrm{B} \Lambda$. Notice that for any $v \in V$ it holds

$$\langle \mathcal{B} v, v\rangle = (\mathrm{B}\Lambda v, \Lambda v)_U \geqslant \underline{c}_{\mathrm{B}} \|\Lambda v\|_U^2 \geqslant \frac{\underline{c}_{\mathrm{B}}}{C_\Lambda^2} \|v\|_{\mathcal{V}}^2. \tag{4.15}$$

If $\{w_i\}$ is a system of linearly independent functions in V, then $\{g_i\}$ form a system of linearly independent functions in $\mathcal{V}$. Indeed, assume the opposite, i.e., $\sum\limits_{i=1}^{m} \zeta_i g_i = \sum\limits_{i=1}^{m} \zeta_i \mathcal{B} w_i = 0$ for some $\zeta_1, \zeta_2, ..., \zeta_m$, which are not all equal to zero. Then for $\bar{v} := \sum\limits_{i=1}^{m} \zeta_i w_i \neq 0$ we have $\mathcal{B}\bar{v} = 0$ and

$$(\mathrm{B}\Lambda\bar{v}, \Lambda\bar{v})_U = \sum_{i,j=1}^{m} (\mathrm{B}\Lambda w_i, \Lambda w_j)_U \zeta_i \zeta_j = \sum_{i=1}^{m} \zeta_i^2 > 0.$$

We arrive at a contradiction with (4.15), which shows that the assumption on g_i was not correct.

Henceforth, T denotes the Gram matrix, whose entries are defined by the relations $T_{ij} = (g_i, g_j)_{\mathcal{V}}$. This matrix is non degenerate and has the inverse matrix T^{-1}. Next, let

$$\begin{aligned} &b \in \mathbb{R}^q,\ b = \{b_k\},\ b_k := (\mathbb{R}, \psi_k)_{\mathcal{V}}, \\ &r \in \mathbb{R}^m,\ r = \{r_i\},\ r_i := (\mathbb{R}, g_i)_{\mathcal{V}}, \\ &S = \{s_{ki}\} \in \mathbb{M}^{q\times m},\ s_{ki} = (\psi_k, g_i)_{\mathcal{V}}, \end{aligned}$$

and $D := ST^{-1}S^T \in \mathbb{M}^{q\times q}$ be a nondegenerate matrix. Here ψ_k, $k = 1, 2, ..., q$, $q < m$ are linearly independent functions forming the set Ψ (cf. (2.6)). We define the set

$$V_m^{\Psi} = \{\varphi \in V_m \mid (\varphi, \psi_k)_{\mathcal{V}} = (\mathbb{R}, \psi_k)_{\mathcal{V}}, \qquad k = 1, 2, ..., q.\},$$

which is a subspace of V_m.

Lemma 2. *For any $e \in V$, $\mathbb{R} \in \mathcal{V}$, and $\mu > 0$, it holds*

$$|\langle \mathbb{R}, e\rangle| \leqslant \frac{\mu}{2}\|\Lambda e\|_{\mathrm{B}}^2 + \frac{1}{2\mu}\left(\|\Lambda u_{\mathrm{R},m}\|_{\mathrm{B}}^2 + \frac{C_{\Psi,\Lambda}^2}{\underline{c}_{\mathrm{B}}}\Big(\|\mathbb{R}\|_{\mathcal{V}}^2 - \Sigma(r, z)\Big)\right), \tag{4.16}$$

where $u_{\mathrm{R},m}$ is defined by (4.10),

$$\Sigma(r, z) := T^{-1} r \cdot r - D^{-1} z \cdot z,$$

$C_{\Psi,\Lambda}$ is a constant in (2.6)*, g_i are defined by* (4.14)*, and $z := b - ST^{-1}r$.*

Proof. In view of (4.11), we have

$$0 = (\mathrm{B}\Lambda(e_m - e), \Lambda w_i)_U = (e - e_m, g_i)_{\mathcal{V}} \quad i = 1, 2, \ldots, m.$$

Therefore, for any $\varphi = \sum\limits_{i=1}^{m} \kappa_i g_i$ it holds

$$(\mathbb{R}, e)_{\mathcal{V}} = (\mathbb{R}, e_m)_{\mathcal{V}} + (\mathbb{R}, e - e_m)_{\mathcal{V}} \\ = (\mathrm{B}\Lambda u_{\mathrm{R},m}, \Lambda e_m)_U + (\mathbb{R} - \varphi, e - e_m)_{\mathcal{V}}. \quad (4.17)$$

The function $\varphi \in V_m$ is at our disposal. We set $\varphi = \bar{\varphi} \in V_m^{\Psi}$, where

$$\|\mathbb{R} - \bar{\varphi}\|_{\mathcal{V}} = \min_{\varphi \in \in V_m^{\Psi}} \|\mathbb{R} - \varphi\|_{\mathcal{V}}. \quad (4.18)$$

Problem (4.18), has another equivalent form

$$\min_{\varphi \in V_m^{\Psi}} \|\mathbb{R} - \varphi\|_{\mathcal{V}}^2 = \min_{\varphi \in V_m} \max_{\xi_k \in \mathbb{R}} \left\{ \|\mathbb{R} - \varphi\|_{\mathcal{V}}^2 + \sum_{k=1}^{q} \xi_k (R - \varphi, \psi_k)_{\mathcal{V}} \right\} \\ = \|\mathbb{R}\|_{\mathcal{V}}^2 - T^{-1} r \cdot r + D^{-1} z \cdot z.$$

Now, we estimate the last term in (4.17) using (4.18) and (2.6)

$$|(\mathbb{R} - \varphi, e - e_m)_{\mathcal{V}}| = |(\mathbb{R} - \varphi, e - e_m - \psi)_{\mathcal{V}}| \\ \leqslant \min_{\varphi \in \mathcal{V}_m^{\Psi}} \|\mathbb{R} - \varphi\|_{\mathcal{V}} \inf_{\psi \in \Psi} \|e - e_m - \psi\|_{\mathcal{V}} \\ \leqslant \left(\|\mathbb{R}\|_{\mathcal{V}}^2 - T^{-1} r \cdot r + D^{-1} z \cdot z \right) C_{\Psi,\Lambda} \|\Lambda(e - e_m)\|_U.$$

Hence (4.17) implies the estimate

$$|(\mathbb{R}, e)_{\mathcal{V}}| \leqslant \|\Lambda u_{\mathrm{R},m}\|_{\mathrm{B}} \|\Lambda e_m\|_{\mathrm{B}} \\ + \frac{C_{\Psi,\Lambda}}{\sqrt{\underline{c}_B}} \left(\|\mathbb{R}\|_{\mathcal{V}}^2 - T^{-1} r \cdot r + D^{-1} z \cdot z \right) \|\Lambda e_m^{\perp}\|_B \\ \leqslant \left(\|\Lambda u_{\mathrm{R},m}\|_{\mathrm{B}}^2 + \frac{C_{\Psi,\Lambda}}{\underline{c}_{\mathrm{B}}} \left(\|\mathbb{R}\|_{\mathcal{V}}^2 - T^{-1} r \cdot r + D^{-1} z \cdot z \right) \right)^{1/2} \|\Lambda e\|_{\mathrm{B}} \quad (4.19)$$

Now (4.16) follows from (4.19) and Young's inequality. □

Remark 2. If we do not impose additional orthogonality conditions and use V_m instead of V_m^{Ψ}, then the constant in (2.5) replaces $G_{\Psi,\Lambda}$ and (4.16) has a simpler form

$$|\langle \mathbb{R}, e \rangle| \leqslant \frac{1}{2\mu} \left(\|\Lambda u_{\mathrm{R},m}\|_{\mathrm{B}}^2 + \frac{C_{\Lambda}^2}{\underline{c}_{\mathrm{B}}} \left(\|\mathbb{R}\|_{\mathcal{V}}^2 - \sum_{i,j=1}^{m} T_{ij}^{-1} r_i r_j \right) \right) + \frac{\mu}{2} \|\Lambda e\|_{\mathrm{B}}^2. \quad (4.20)$$

This estimate has a clear meaning. Let $\mathcal{V}_m = \text{span}\{g_1, g_2, ..., g_m\}$ and

$$\mathcal{V}_m^\perp := \Big\{ v \in \mathcal{V} \mid (v, g_i)_\mathcal{V} = 0, i = 1, 2, ..., m \Big\}.$$

In this case, the coefficients κ_i are defined by the minimisation problem $\min\limits_{\kappa_i} \|\mathbb{R} - \sum\limits_{i=1}^m \kappa_i g_i\|_\mathcal{V}^2$, so that the function $\mathbb{R}_m = \sum\limits_{i=1}^m \kappa_i g_i$ is the orthogonal projection of $\mathbb{R}$ to $\mathcal{V}_m$. Hence $\mathbb{R} = \mathbb{R}_m + \mathbb{R}_m^\perp$, where $\mathbb{R}_m^\perp \in \mathcal{V}_m^\perp$ and

$$\|\mathbb{R}\|_\mathcal{V}^2 - \sum_{i,j=1}^m T_{ij}^{-1} r_i r_j = \|\mathbb{R}_m^\perp\|_\mathcal{V}^2. \tag{4.21}$$

By (4.21) we rewrite (4.20) as follows:

$$|\langle\, \mathbb{R}, e \,\rangle| \leqslant \frac{1}{2\mu}\left(\|\Lambda u_{\mathrm{R},m}\|_\mathrm{B}^2 + \frac{C_\Lambda^2}{\underline{c}_\mathrm{B}} \|\mathbb{R}_m^\perp\|_\mathcal{V}^2 \right) + \frac{\mu}{2}\|\Lambda e\|_\mathrm{B}^2.$$

Unlike (4.3), this estimate includes only a part of $\mathbb{R}$ that belongs to the orthogonal complement $\mathcal{V}_m^\perp$. The larger is m the smaller is $\|\mathbb{R}_m^\perp\|_\mathcal{V}$. Theoretically, if the subspaces $\{\mathcal{V}_m\}$ are limit dense in $\mathcal{V}$ (what will be if $\{V_m\}$ are limit dense in V) then $\|\mathbb{R}_m^\perp\|_\mathcal{V} \to 0$ as $m \to \infty$. Certainly, in practice we limit ourselves to some finite m.

4.3. Problem (3.1)–(3.2). The estimate (4.20) holds for any $e \in V$ regardless of the origin of this error function. Consider the case, where $\mathrm{B} = A$, $\ell \in \mathcal{V}$, $e = v - u$, $v \in V$, and u solves the problem

$$(A\Lambda u, \Lambda w)_U + (\ell, w)_\mathcal{V} = 0. \tag{4.22}$$

In this case, the term $(\mathbb{R}, e_m)_\mathcal{V}$ can be explicitly computed. Error component $e_m = \sum\limits_{i=1}^m \alpha_i w_i$ is defined by the coefficients α_i, which can be found by the orthogonality relation

$$\begin{aligned} 0 = (A\Lambda(e_m - e), \Lambda w_j)_U &= (A\Lambda(e_m - v + u), \Lambda w_j)_U \\ &= \sum_{i=1}^m \alpha_i (A\Lambda w_i, \Lambda w_j)_U - (A\Lambda v, \Lambda w_j)_U - (\ell, w_j)_\mathcal{V}. \end{aligned}$$

Recalling (4.13), we find that

$$\alpha_i = (\ell, w_i)_\mathcal{V} + (A\Lambda v, \Lambda w_i)_U. \tag{4.23}$$

Thus,

$$\|\Lambda e_m\|_A^2 = \sum_{i=1}^m \alpha_i^2, \qquad (\mathbb{R}, e_m)_{\mathcal{V}} = \sum_{i=1}^m \alpha_i \rho_i, \quad \rho_i := (\mathbb{R}, w_i)_{\mathcal{V}},$$

and instead of (4.17) we have

$$(\mathbb{R}, e)_{\mathcal{V}} = (\mathbb{R}, e_m)_{\mathcal{V}} + (\mathbb{R}, e - e_m)_{\mathcal{V}} = \sum_{i=1}^m \alpha_i \rho_i + (\mathbb{R} + \varphi, e - e_m)_{\mathcal{V}},$$

where $\varphi = \sum\limits_{i=1}^m \kappa_i g_i \in \mathcal{V}_m$. We estimate the last term and derive the estimates

$$\begin{aligned}(\mathbb{R}, e)_{\mathcal{V}} &\leqslant \sum_{i=1}^m \alpha_i \rho_i + \Big(\|\mathbb{R}\|_{\mathcal{V}}^2 - T^{-1} r \cdot r\Big)^{1/2} \|e_m^{\perp}\|_{\mathcal{V}} \\ &\leqslant \sum_{i=1}^m \alpha_i \rho_i + \frac{C_\Lambda}{\sqrt{\underline{c}_A}} \Big(\|\mathbb{R}\|_{\mathcal{V}}^2 - T^{-1} r \cdot r\Big)^{1/2} \|\Lambda e\|_{\mathcal{V}} \end{aligned} \tag{4.24}$$

and (for $\mu > 0$)

$$(\mathbb{R}, e)_{\mathcal{V}} \leqslant \sum_{i=1}^m \alpha_i \rho_i + \frac{C_\Lambda^2}{2\mu \underline{c}_A} \Big(\|\mathbb{R}\|_{\mathcal{V}}^2 - T^{-1} r \cdot r\Big) + \frac{\mu}{2} \|\Lambda e\|_A^2. \tag{4.25}$$

The estimate (4.25) is sharper than (4.20) because the first quantity in the right hand side is equal to $(\mathbb{R}, e_m)_{\mathcal{V}}$. In (4.16), it is estimated from above by means of $u_{\mathrm{R},m}$.

4.4. Decomposition of $\langle \mathbb{R}, e \rangle$ by a set of eigenfunctions. Let the set of functions w_i, $i = 1, 2, ..., m$ be formed by eigenfunctions of the operator $\Lambda^* \Lambda$. In this case,

$$\begin{aligned} &(\Lambda w_i, \Lambda w_j)_U = \langle \Lambda^* \Lambda w_i, w_j \rangle = \lambda_i (w_i, w_j)_{\mathcal{V}} = 0 \quad i \neq j, \\ &(w_i, w_i)_{\mathcal{V}} = \frac{1}{\lambda_i}, \qquad \|\Lambda w_i\|_U^2 = 1. \end{aligned}$$

The subspaces

$$V_m = \operatorname{span}\big\{w_1, w_2, ..., w_m\big\} \text{ and } V_m^{\perp} = \big\{v \in V \ \mid \ (v, w_i)_{\mathcal{V}} = 0, \ i = 1, 2, ..., m\big\}$$

create orthogonal decomposition $V = V_m \oplus V_m^{\perp}$ with respect to the product $(\cdot, \cdot)_{\mathcal{V}}$.

Lemma 3. *For any $\mathbb{R} \in \mathcal{V}$, $e \in V$, and $\mu > 0$ it holds*

$$|(\mathbb{R}, e)_{\mathcal{V}}| \leqslant \frac{1}{2\mu}\left(\sum_{i=1}^{m} \rho_i^2 + \frac{1}{\lambda_{m+1}}\|\mathbb{R}_m^{\perp}\|_{\mathcal{V}}^2\right) + \frac{\mu}{2}\|\Lambda e\|_U^2, \tag{4.26}$$

where

$$\rho_i = (\mathbb{R}, w_i)_{\mathcal{V}} \quad \textit{and} \quad \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}}^2 = \|\mathbb{R}\|_{\mathcal{V}}^2 - \sum_{i=1}^{m} \lambda_i \rho_i^2. \tag{4.27}$$

Proof. As before, we use decomposition of the error $e = e_m + e_m^{\perp}$, which implies decomposition of the key term

$$(\mathbb{R}, e)_{\mathcal{V}} = (\mathbb{R}, e_m)_{\mathcal{V}} + (\mathbb{R}, e_m^{\perp})_{\mathcal{V}}. \tag{4.28}$$

Since $(\Lambda e_m, \Lambda e_m^{\perp})_U = 0$, we see that in this case errors satisfy a Pythagorean type identity in the space U

$$\|\Lambda e\|_U^2 = \|\Lambda e_m\|_U^2 + \|\Lambda e_m^{\perp}\|_U^2. \tag{4.29}$$

To estimate the first term of (4.28), we use the auxiliary finite dimensional problem: find $u_{\mathrm{R},m} = \sum\limits_{i=1}^{m} \beta_i w_i$ such that

$$(\Lambda u_{\mathrm{R},m}, \Lambda w) = (\mathbb{R}, w)_{\mathcal{V}}, \qquad \forall w \in V_m. \tag{4.30}$$

The coefficients β_i are defined by the system

$$\sum_{i=1}^{m} \beta_i (\Lambda w_i, \Lambda w_j) = (\mathbb{R}, w_j)_{\mathcal{V}} = \rho_j, \quad j = 1, 2, ..., m.$$

It is easy to see that $\beta_i = \rho_i$. Hence

$$\|\Lambda u_{\mathrm{R},m}\|_U^2 = \sum_{i=1}^{m} \beta_i^2 \|\Lambda w_j\|_U^2 = \sum_{i=1}^{m} \rho_i^2$$

and by (4.30) we find that

$$|\langle \mathbb{R}, e_m \rangle| \ = \ |(\Lambda u_{\mathrm{R},m}, \Lambda e_m)| \ \leqslant \|\Lambda u_{\mathrm{R},m}\|_U \|\Lambda e_m\|_U. \tag{4.31}$$

Consider another part of (4.28). We have

$$(\mathbb{R}, e_m^{\perp})_{\mathcal{V}} = (\mathbb{R} - \sum_{i=1}^{m} \kappa_i w_i, e_m^{\perp})_{\mathcal{V}}.$$

We set $\kappa_i = \lambda_i \rho_i$ and find that

$$(\mathbb{R}, e_m^{\perp})_{\mathcal{V}} \leqslant \Big(\|\mathbb{R}\|^2_{\mathcal{V}} - \sum_{i=1}^{m} \lambda_i \rho_i^2\Big)^{1/2} \|e_m^{\perp}\|_{\mathcal{V}} = \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}} \|e_m^{\perp}\|_{\mathcal{V}}. \tag{4.32}$$

Notice that $e_m^{\perp} \in V_m^{\perp}$ and, therefore,

$$\|e_m^{\perp}\|_{\mathcal{V}} \leqslant \frac{1}{\sqrt{\lambda_{m+1}}} \|\Lambda e_m^{\perp}\|_U. \tag{4.33}$$

By (4.32) and (4.33) we obtain

$$\big|\langle \mathbb{R}, e_m^{\perp}\rangle\big| \leqslant \frac{1}{\sqrt{\lambda_{m+1}}} \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}} \|\Lambda e_m^{\perp}\|_U. \tag{4.34}$$

From (4.28), (4.31), and (4.34), it follows that

$$\begin{aligned} |\langle \mathbb{R}, e\rangle| &\leqslant \left(\sum_{i=1}^{m} \rho_i^2\right)^{1/2} \|\Lambda e_m\|_U + \frac{1}{\sqrt{\lambda_{m+1}}} \|\mathbb{R}_m^{\perp}\|^2_{\mathcal{V}} \|\Lambda e_m^{\perp}\|_U \\ &\leqslant \left(\sum_{i=1}^{m} \rho_i^2 + \frac{1}{\lambda_{m+1}} \|\mathbb{R}_m^{\perp}\|^2_{\mathcal{V}}\right)^{1/2} \|\Lambda e\|_U. \end{aligned} \tag{4.35}$$

We arrive at (4.17) by applying Young's inequality to the right hand side of (4.35). □

Remark 3. Let $\mathrm{B} = \mathbb{1}$ and, consequently, $\underline{c}_{\mathrm{B}} = 1$. Compare (4.17) and (4.26). Since

$$\frac{1}{C_\Lambda} = \inf_{w \in V} \frac{\|\Lambda w\|_U}{\|w\|_{\mathcal{V}}} = \lambda_1^{1/2},$$

we have $C_\Lambda^2 = \frac{1}{\lambda_1} \geqslant \frac{1}{\lambda_{m+1}}$. The first term in round brackets is the same in both relations (it represents $\|\Lambda u_{\mathrm{R},m}\|^2_U$). Hence (4.26) is sharper than (4.17).

4.5. Particular case. Now we focus attention on a special, but important case where the error function e is generated by the problem (3.1)–(3.2) with $A = \mathbb{1}$ so that u satisfies the equation

$$\Lambda^* \Lambda u + \ell = 0. \tag{4.36}$$

This equation is an abstract form of elliptic equations associated with self–adjoint operators, which are used in natural sciences.

Notice that the first m coefficients γ_i in the representation $u = \sum_{i=1}^{\infty} \gamma_i w_i$ can be defined by solving a system of linear simultaneous equations

$$\sum_{i=1}^{m} \gamma_i (\Lambda w_i, \Lambda w_j)_U + (\ell, w_j) = 0, \quad j = 1, 2, ..., m, \tag{4.37}$$

wherefrom $\gamma_i = -\ell_i$, $\ell_i := (\ell, w_i)_{\mathcal{V}}$. We use the identity (3.3)

$$\|e^*\|_U^2 + \|\Lambda e\|_U^2 = \|y - \Lambda v\|_U^2 + 2(\mathbb{R}, e)_{\mathcal{V}}, \quad \mathbb{R} := \Lambda^* y + \ell. \tag{4.38}$$

To estimate the last term we decompose the error: $e = e_m + e_m^{\perp}$. Let $\zeta_i := \lambda_i (v, w_i)_{\mathcal{V}}$, then the first summand of this decomposition is computable:

$$e_m = v_m - u_m = \sum_{i=1}^{m} (\zeta_i - \gamma_i)\, w_i = \sum_{i=1}^{m} (\zeta_i + \ell_i) w_i.$$

We have

$$(\mathbb{R}, e)_{\mathcal{V}} = (\mathbb{R}, e_m)_{\mathcal{V}} + (\mathbb{R}, e_m^{\perp})_{\mathcal{V}} = \Sigma_m + (\mathbb{R}_m^{\perp}, e_m^{\perp})_{\mathcal{V}},$$

where

$$\Sigma_m := \sum_{i=1}^{m} \rho_i\, (\zeta_i - \gamma_i) = \sum_{i=1}^{m} (\zeta_i + \ell_i)\, \rho_i.$$

By (4.34), we conclude that

$$\begin{aligned}(\mathbb{R}, e)_{\mathcal{V}} &\leqslant \Sigma_m + \frac{1}{\sqrt{\lambda_{m+1}}} \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}}\, \|\Lambda e\|_U \\ &\leqslant \Sigma_m + \frac{1}{2\mu\lambda_{m+1}} \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}}^2 + \frac{\mu}{2} \|\Lambda e\|_U^2 \end{aligned} \tag{4.39}$$

and

$$(\mathbb{R}, e)_{\mathcal{V}} \geqslant \Sigma_m - \frac{1}{\sqrt{\lambda_{m+1}}} \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}}\, \|\Lambda e\|_U.$$

Thus, (4.38) yields two sided bounds

$$(1 - \mu)\|\Lambda e\|_U^2 + \|e^*\|_U^2 \leqslant \|y - \Lambda v\|_U^2 + 2\Sigma_m + \frac{1}{\mu\lambda_{m+1}} \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}}^2, \tag{4.40}$$

$$(1 + \mu)\|\Lambda e\|_U^2 + \|e^*\|_U^2 \geqslant \|y - \Lambda v\|_U^2 + 2\Sigma_m - \frac{1}{\mu\lambda_{m+1}} \|\mathbb{R}_m^{\perp}\|_{\mathcal{V}}^2, \tag{4.41}$$

where $\mu \in (0, 1]$ in (4.40) and $\mu > 0$ in (4.41).

4.6. Sharpness of the estimate. Possible overestimation in (4.40) and underestimation in (4.41) is associated exclusively with the last term, and depends on the quantity

$$E_m := \frac{1}{\lambda_{m+1}}\Big(\|\mathbb{R}\|^2_{\mathcal{V}} - \sum_{i=1}^{m}\lambda_i\rho_i^2\Big).$$

Since

$$\lambda_{m+1}\to\infty \quad \text{and} \quad \sum_{i=1}^{m}\lambda_i\rho_i^2 \to \|\mathbb{R}\|^2_{\mathcal{V}} \text{ as } m\to\infty,$$

this term tends to zero. Hence the estimates (4.40) and (4.41) become sharper as m grows and converge to the exact value from above and below. Consider one case where this fact very easy to observe. Let $y = \Lambda v$. Then $e^* = \Lambda e$ and (4.40) reads

$$(1-\frac{\mu}{2})\|\Lambda e\|^2_U \leqslant \Sigma_m\left(1+\frac{\kappa}{\mu}\right), \quad \kappa = \frac{E_m}{2\Sigma_m}$$

Set $\mu = \sqrt{2\kappa}$. Then

$$\|\Lambda e\|^2_U \leqslant \Sigma_m \frac{1+\sqrt{\kappa/2}}{1-\sqrt{\kappa/2}} \tag{4.42}$$

Notice that

$$\rho_i = (\Lambda^*\Lambda v + \ell, w_i) = (\sum_{j=1}^{\infty}\zeta_j\Lambda w_j, \Lambda w_i) + (\ell, w_i) = \zeta_i + \ell_i.$$

Therefore,

$$\Sigma_m = \sum_{i=1}^{m}(\zeta_i+\ell_i)\rho_i = \sum_{i=1}^{m}\rho_i^2 \quad \text{and} \quad E_m = \frac{1}{\lambda_{m+1}}\sum_{i=m+1}^{\infty}\lambda_i\rho_i^2.$$

Since

$$\|\Lambda e\|^2_U = \Big(\sum_{i=1}^{\infty}(\zeta_i+\ell_i)\Lambda w_i, \sum_{i=1}^{\infty}(\zeta_i+\ell_i)\Lambda w_i\Big) = \sum_{i=1}^{\infty}(\zeta_i+\ell_i)^2 = \sum_{i=1}^{\infty}\rho_i^2,$$

we see that $E_m \to 0$ and $\Sigma_m \to \|\Lambda e\|^2$, and κ monotonically tends to zero as $m \to +\infty$. Hence the left hand side of (4.42) tends to the right hand one.

4.7. Extension of the applicability area. Above discussed method is based on the knowledge of m eigenfunctions. This fact imposes restrictions on the shape of Ω. We can partially bypass them if V consists of the functions satisfying homogeneous boundary conditions on $\partial\Omega$.

Let $\Omega \subset \widehat{\Omega}$, where $\widehat{\Omega}$ is a "simple" domain for which the eigenfunctions $\widehat{w}_i$ are known. We extend u by setting $u = 0$ in $\widehat{\Omega} \setminus \Omega$ and denote it by $\widehat{u}$. Analogously $\widehat{v}$ and $\widehat{\mathbb{R}}$ are the extensions of v and $\mathbb{R}$ by zero. Then,

$$\widehat{e} := \widehat{v} - \widehat{u} = \begin{cases} e & \text{in } \Omega \\ 0 & \text{in } \widehat{\Omega} \setminus \Omega \end{cases}$$

and

$$(\mathbb{R}, e)_{\mathcal{V}(\Omega)} = (\widehat{\mathbb{R}}, \widehat{e})_{\mathcal{V}(\widehat{\Omega})}. \tag{4.43}$$

If the operator Λ admits analogous extension, which preseves the norm

$$\|\Lambda e\|_{U(\Omega)} = \|\Lambda \widehat{e}\|_{U(\widehat{\Omega})}, \tag{4.44}$$

then by Lemma 3 we have

$$\left|(\widehat{\mathbb{R}}, \widehat{e})_{\mathcal{V}(\widehat{\Omega})}\right| \leqslant \frac{1}{2\mu}\left(\sum_{i=1}^{m} \widehat{\rho}_i^2 + \frac{1}{\widehat{\lambda}_{m+1}}\left(\|\widehat{\mathbb{R}}\|^2_{\mathcal{V}(\widehat{\Omega})} - \sum_{i=1}^{m} \widehat{\lambda}_i \widehat{\rho}_i^2\right)\right) + \frac{\mu}{2}\|\Lambda \widehat{e}\|^2_{U(\widehat{\Omega})}. \tag{4.45}$$

where $\widehat{\rho}_i = (\widehat{\mathbb{R}}, \widehat{w}_i)_{\mathcal{V}(\widehat{\Omega})}$. From (4.43), (4.44), and (4.45) it follows that

$$\left|(\mathbb{R}, e)_{\mathcal{V}(\Omega)}\right| \leqslant \frac{1}{2\mu}\left(\sum_{i=1}^{m} \widehat{\rho}_i^2 + \frac{1}{\widehat{\lambda}_{m+1}}\left(\|\mathbb{R}\|^2_{\mathcal{V}(\Omega)} - \sum_{i=1}^{m} \widehat{\lambda}_i \widehat{\rho}_i^2\right)\right) + \frac{\mu}{2}\|\Lambda e\|^2_{U(\Omega)}. \tag{4.46}$$

Thus, we get an upper bound using known eigenfunctions associated with $\widehat{\Omega}$.

Also, it is worth noting that the method outlined in Sec. 4.5 can be also used to improve the approximation v. With no additional expenditures we find the function

$$v_m^+ := v - \sum_{i=1}^{m} \frac{1}{\lambda_i}(\lambda_i \zeta_i + \ell_i) w_i,$$

which is as a better approximation of u than v_m. Indeed,

$$v_m^+ = v - e_m = v_m + v_m^\perp - v_m + u_m = v_m^\perp + u - u_m^\perp = u + e_m^\perp.$$

Therefore, the corresponding error $e_m^+ = v_m^+ - u = e_m^\perp$ has zero projection on V_m and $(v, w_j)_{\mathcal{V}} = (v_m^+, w_j)_{\mathcal{V}}$ for any $j \geqslant m+1$. Recalling (4.29), we conclude that the error of v_m^+ is smaller than the error of v.

4.8. Generalizations. It may seem that the whole topic about getting estimates (4.40) and (4.41) for the problem (4.36) is rather special, but this is not the case. For any of the mathematical models considered in Sec. 2 (and for many others), fully reliable error control problem can be reduced to the case studied in Sec. 4.5. Indeed, we can replace (4.12) by

$$\langle \mathbb{R}, e \rangle = (\mathrm{B}\Lambda v_{\mathrm{R}}, \Lambda e)_U + (\mathrm{B}\Lambda(u_{\mathrm{R}} - v_{\mathrm{R}}), \Lambda e)_U, \tag{4.47}$$

where v_{R} is an approximation of u_{R} defined by (4.8) and $e_{\mathrm{R}} := v_{\mathrm{R}} - u_{\mathrm{R}}$. Then

$$\langle \mathbb{R}, e \rangle \leqslant (\|\Lambda v_{\mathrm{R}}\|_U + \|\Lambda e_{\mathrm{R}}\|_{\mathrm{B}})\|\Lambda e\|_B.$$

and

$$\langle \mathbb{R}, e \rangle \leqslant \frac{1}{\mu}\Big(\|\Lambda v_{\mathrm{R}}\|_U^2 + \|\Lambda e_{\mathrm{R}}\|_B^2\Big) + \frac{\mu}{2}\|\Lambda e\|_B^2, \quad \forall \mu > 0. \tag{4.48}$$

This estimate has the same structure as (4.20) and (4.26). It yields guaranteed and computable error estimates provided that the first term in the right hand side of (4.48) is computable. This goal is achieved if we solve the linear problem (4.6) numerically and find a sharp computable bound for the norm of the corresponding error e_{R}. We are free to choose a method of finding v_{R} and it is not required that it is the exact solution (Galerkin approximation) of (4.8). Thus, (4.48) shows *principal ability to control the accuracy of approximations to various problems using an auxiliary linear problem only.*

To estimate e_{R} we set $\mathrm{B} = \mathbb{1}$ and use the method considered in Sec. 4.5. The auxiliary problem (4.8) reads

$$\Lambda^* \Lambda u_{\mathrm{R}} + \mathbb{R} = 0. \tag{4.49}$$

Let v_{R} and y_{R} be approximations of u_{R} and $p_{\mathrm{R}} := \Lambda u_{\mathrm{R}}$, respectively. Then

$$e_{\mathrm{R}} = v_{\mathrm{R}} - u_{\mathrm{R}}, \quad e_{\mathrm{R}}^* = y_{\mathrm{R}} - p_{\mathrm{R}}, \quad \text{and} \quad \Lambda^* e_{\mathrm{R}}^* = \Lambda^* y_{\mathrm{R}} + \mathbb{R} =: \mathfrak{R}(y_{\mathrm{R}}).$$

For (4.49) we use (3.7) with $z = 0$ and $A = \mathbb{1}$. It reads

$$(1-\gamma\epsilon)\|\Lambda e\|_U^2 + \big(1 - \tfrac{\gamma}{\epsilon}\big)\,\|e^*\|_U^2 \leqslant \|y - \Lambda v\|_U^2 + 2(1-\gamma)(e^*, \Lambda e), \tag{4.50}$$

where γ and ϵ are parameters satisfying the conditions

$$\gamma \geqslant 0, \quad \epsilon > 0, \quad \gamma\epsilon \leqslant 1, \quad \text{and } \epsilon \geqslant \gamma.$$

If $\gamma = \epsilon$, $v = v_{\mathrm{R}}$, $y = y_{\mathrm{R}}$, and $\mathfrak{R}(y_{\mathrm{R}}) \in \mathcal{V}$, then (4.50) has the form

$$(1-\gamma^2)\|\Lambda e_{\mathrm{R}}\|_U^2 \;\leqslant\; \|y_{\mathrm{R}} - \Lambda v_{\mathrm{R}}\|_U^2 + 2(1-\gamma)(\mathfrak{R}(y_{\mathrm{R}}), e_{\mathrm{R}})_{\mathcal{V}}.$$

To estimate the term $(\mathfrak{R}(y_{\mathrm{R}}), e_{\mathrm{R}})_{\mathcal{V}}$, we use (4.39), which reads

$$(\mathfrak{R}(y_{\mathrm{R}}), e_{\mathrm{R}})_{\mathcal{V}} \leqslant \Sigma_m + \frac{1}{2\nu\lambda_{m+1}}\|\mathfrak{R}(y_{\mathrm{R}})_m^{\perp}\|_{\mathcal{V}}^2 + \frac{\nu}{2}\|\Lambda e_{\mathrm{R}}\|_U^2, \quad \forall \nu > 0.$$

Combining the last two inequalities, we obtain the estimate

$$\begin{aligned}(1-\nu+\gamma\nu-\gamma^2)\|\Lambda e_{\mathrm{R}}\|_U^2 \leqslant{}& \|y_{\mathrm{R}} - \Lambda v_{\mathrm{R}}\|_U^2 \\ &+ 2(1-\gamma)\Sigma_m + \frac{(1-\gamma)}{\nu\lambda_{m+1}}\|\mathfrak{R}(y_{\mathrm{R}})_m^{\perp}\|_{\mathcal{V}}^2, \end{aligned} \tag{4.51}$$

where

$$\Sigma_m = \sum_{i=1}^{m}\Big((v_{\mathrm{R}}, w_i)_{\mathcal{V}} + \frac{(\mathbb{R}, w_i)_{\mathcal{V}}}{\lambda_i}\Big)(\mathfrak{R}(y_{\mathrm{R}}), w_i)_{\mathcal{V}}$$

and

$$\|\mathfrak{R}(y_{\mathrm{R}})_m^{\perp}\|_{\mathcal{V}}^2 = \|\mathfrak{R}(y_{\mathrm{R}})\|_{\mathcal{V}}^2 - \sum_{i=1}^{m}(\mathfrak{R}(y_{\mathrm{R}}), w_i)_{\mathcal{V}}^2.$$

Let $\nu = \gamma < 1$. Then (4.51) has the form

$$\|\Lambda e_{\mathrm{R}}\|_U^2 \leqslant \frac{\|y_{\mathrm{R}} - \Lambda v_{\mathrm{R}}\|_U^2}{1-\gamma} + 2\Sigma_m + \frac{1}{\gamma\lambda_{m+1}}\|\mathfrak{R}(y_{\mathrm{R}})_m^{\perp}\|_{\mathcal{V}}^2. \tag{4.52}$$

Estimates (4.48) and (4.52) imply an upper bound of $|(\mathbb{R}, e)_{\mathcal{V}}|$. For example, set $\gamma = \frac{1}{2}$ in (4.52). Then, for any $\mu > 0$ we have

$$\begin{aligned}|(\mathbb{R}, e)_{\mathcal{V}}| \;\leqslant{}& \frac{1}{\mu}\|\Lambda v_{\mathrm{R}}\|_U^2 + \frac{\mu}{2}\|\Lambda e\|_U^2 \\ &+ \frac{2}{\mu}\Big(\Sigma_m + \|y_{\mathrm{R}} - \Lambda v_{\mathrm{R}}\|_U^2 + \frac{1}{\lambda_{m+1}}\|\mathfrak{R}(y_{\mathrm{R}})_m^{\perp}\|_{\mathcal{V}}^2\Big).\end{aligned}$$

All terms in the right hand side of this estimate (except $\|\Lambda e\|^2$) depend only on approximate solutions of (4.49) and known data.

§5. A posteriori error estimates

In the literature, various a posteriori error estimates are mainly studied in the context of adaptive computational methods. There are known several types of a posteriori indicators (residual, hierarchical, goal–oriented, post-processing, etc.), which are relatively cheap and usually suggest a correct way of changing meshes (or other parameters of approximations) in order to get the best approximate solution at the next iteration. Theses methods exploit specific features of approximations (e.g., Galerkin orthogonality) and properties of exact solutions (e.g., additional regularity) to construct simply computable indicators of errors. Typically, they are adapted to a particular problem or numerical method. The reader will find the corresponding theory and numerous examples in [1, 3, 6–8, 10, 26, 27] and many other publications cited therein.

Estimates considered in this section follow another concept. They are derived from a posteriori error identities that hold for *all* deviations from the exact solutions. The only one requirement is that they must belong to the same functional class as the generalised solution of the problem under consideration. We show that above derived estimates of the term $\langle \mathbb{R}, e\rangle$ imply guaranteed and fully computable error bounds for various boundary value problems, which are valid for the same wide class of deviations. The identities and estimates are obtained by general methods of functional analysis and theory of boundary value problems without attracting any additional information (e.g., on properties of approximations or numerical methods). Therefore, they are often called a posteriori estimates/identities of the *functional type*. They possess maximal universality and can be also applied to analysis of modeling errors [19].

5.1. Problem (3.4). We use (3.8) in the form

$$(1-\gamma^2)\|\Lambda e\|^2_A \leqslant \|y - A\Lambda v - z\|^2_{A^{-1}} + 2(1-\gamma)(\mathbb{R}(y), e)_{\mathcal{V}} \quad (5.1)$$

and apply (4.25) to the last term. We find that

$$|(\mathbb{R}(y), e)_{\mathcal{V}}| \leqslant \sum_{i=1}^{m} \alpha_i \rho_i + \frac{C^2_{\Psi,\Lambda}}{2\underline{c}_A \mu}\Big(\|\mathbb{R}\|^2_{\mathcal{V}} - \Sigma(r, z)\Big) + \frac{\mu}{2}\|\Lambda e^{\perp}_m\|^2_A.$$

In the case considered, $\mathrm{B} = A$ and (4.13) reads

$$(A\Lambda w_i, \Lambda w_j)_U = \delta_{ij} \quad i, j = 1, 2, ..., m.$$

Therefore,

$$\|\Lambda e\|_A^2 = \|\Lambda e_m\|_A^2 + \|\Lambda e_m^\perp\|_A^2, \tag{5.2}$$

where $\|\Lambda e_m\|_A^2 = \sum\limits_{i=1}^m \alpha_i^2$. From (5.1) and (5.2), it follows that

$$\begin{aligned}(1-\gamma^2)\|\Lambda e_m^\perp\|_A^2 &\leqslant \|y - A\Lambda v - z\|_{A^{-1}}^2 \\ &\quad + 2(1-\gamma)(\mathbb{R}(y), e)_{\mathcal{V}} - (1-\gamma^2)\sum_{i=1}^m \alpha_i^2 \\ &\leqslant \|y - A\Lambda v - z\|_{A^{-1}}^2 - (1-\gamma^2)\sum_{i=1}^m \alpha_i^2 \\ + (1-\gamma)\Big(2\sum_{i=1}^m \alpha_i\rho_i &+ \frac{C_{\Psi,\Lambda}^2}{\underline{c}_A\mu}\Big(\|\mathbb{R}\|_{\mathcal{V}}^2 - \Sigma(r,z)\Big) + \frac{\mu}{\underline{c}_A}\|\Lambda e_m^\perp\|_A^2\Big),\end{aligned}$$

wherefrom we deduce the estimate

$$\begin{aligned}&(1-\gamma)(1+\gamma-\frac{\mu}{\underline{c}_A})\|\Lambda e_m^\perp\|_A^2 \leqslant \|y - A\Lambda v - z\|_{A^{-1}}^2 \\ &+ (1-\gamma)\Big(2\sum_{i=1}^m \alpha_i\rho_i + \frac{C_{\Psi,\Lambda}^2}{\underline{c}_A\mu}\Big(\|\mathbb{R}\|_{\mathcal{V}}^2 - \Sigma(r,z)\Big) - (1+\gamma)\sum_{i=1}^m \alpha_i^2\Big)\end{aligned} \tag{5.3}$$

Set $\mu = \underline{c}_A\gamma$,

$$I_1(y,v,z) := \|y - A\Lambda v - z\|_{A^{-1}} \quad \text{and} \quad I_2(y) := \frac{C_\Lambda}{\sqrt{\underline{c}_A}}\Big(\|\mathbb{R}\|_{\mathcal{V}}^2 - \Sigma(r,z)\Big)^{1/2}.$$

Then we arrive at the estimates

$$\|\Lambda e_m^\perp\|_A^2 \leqslant \frac{1}{1-\gamma}I_1^2(y,v,z)^2 + \frac{1}{\gamma}I_2^2(y) + \sum_{i=1}^m\Big(2\alpha_i\rho_i - (1+\gamma)\alpha_i^2\Big) \tag{5.4}$$

and

$$\|\Lambda e\|_A^2 \leqslant \frac{1}{1-\gamma}I_1^2(y,v,z)^2 + \frac{1}{\gamma}I_2^2(y) + \sum_{i=1}^m\Big(2\alpha_i\rho_i - \gamma\alpha_i^2\Big). \tag{5.5}$$

Setting in (5.5) $\gamma = \frac{I_2}{I_1+I_2}$, we obtain

$$\|\Lambda e\|_A^2 \leqslant \Big(I_1(y,v,z) + I_2(y)\Big)^2 + 2\sum_{i=1}^m \alpha_i\rho_i - \frac{I_2(y)\sum\limits_{i=1}^m \alpha_i^2}{I_1(y,v,z) + I_2(y)}. \tag{5.6}$$

5.2. Convection–diffusion problem (3.11)–(3.12). We set $\mathrm{B} = \mathbb{1}$ and define the auxiliary problem (4.6) as follows: find $u_{\mathrm{R}} \in V := \overset{\circ}{H}{}^1(\Omega)$ such that

$$\int\limits_{\Omega} \nabla u_{\mathrm{R}} \cdot \nabla w dx = \int\limits_{\Omega} \mathbb{R}(v, y)\, w dx \qquad \forall w \in V, \tag{5.7}$$

where $\mathbb{R}(y, v) := \mathrm{div} y - \mathrm{a} \cdot \nabla v + f$.

Let $\{w_i\}$, $i = 1, 2, ..., m$ be a system of functions in V_0 such that

$$\int\limits_{\Omega} \nabla w_i \cdot \nabla w_j dx = \delta_{ij}.$$

In accordance with (4.14) (where $\mathcal{B} = \Delta$), we additionally require that

$$\Delta w_i \in \mathcal{V} := L_2(\Omega).$$

A finite dimensional counterpart of (5.7) is the problem: find $u_{\mathrm{R},m} \in V_m \subset V_0$ such that

$$\int\limits_{\Omega} \nabla u_{\mathrm{R},m} \cdot \nabla w dx = \int\limits_{\Omega} \mathbb{R}(y, v)\, w dx \qquad \forall w \in V_m. \tag{5.8}$$

First, we use (4.20) and (3.13) to get the following error majorant

$$(1-\mu)\|\nabla e\|^2_{\Omega} + \|e^*\|^2_{\Omega} \leqslant \|\nabla v - y\|^2_{\Omega} + \frac{1}{\mu}\|\nabla u_{\mathrm{R},m}\|^2_{\Omega} + \frac{C_F^2}{\mu}\Big(\|\mathbb{R}(y, v)\|^2_{\Omega} - T^{-1} r \cdot r\Big), \tag{5.9}$$

where C_F is a constant in the Friedrichs inequality $\|w\|_{\Omega} \leqslant C_F \|\nabla w\|_{\Omega}$ for any $w \in \overset{\circ}{H}{}^1(\Omega)$,

$$\mu \in (0, 1], \quad r_i = -\int\limits_{\Omega} \mathbb{R}(y, v) \Delta w_i dx, \;\text{ and }\; T_{ij} = \int\limits_{\Omega} \Delta w_i \Delta w_j dx.$$

It is easy to see that the left hand side of (5.9) is fully fully computable.

Also, we have a lower error bound

$$(1+\mu)\|\nabla e\|^2_{\Omega} + \|e^*\|^2_{\Omega} \geqslant \|\nabla v - y\|^2_{\Omega} - \frac{1}{\mu}\|\nabla u_{\mathrm{R},m}\|^2_{\Omega} - \frac{C_F^2}{\mu}\Big(\|\mathbb{R}(y, v)\|^2_{\Omega} - T^{-1} r \cdot r\Big), \tag{5.10}$$

where $\mu > 0$.

If $\{w_i\}$ are the eigenfunctions of Δ, then Lemma 3 can be applied. In this case, we obtain other bounds

$$(1-\mu)\|\nabla e\|^2_\Omega + \|e^*\|^2_\Omega \leqslant \int\limits_\Omega |\nabla v - y|^2 dx + \frac{1}{\mu}\sum_{i=1}^m \rho_i^2 + \frac{1}{\mu\lambda_{m+1}}\Big(\|\mathbb{R}(y,v)\|^2_\Omega - \sum_{i=1}^m \lambda_i\rho_i^2\Big) \tag{5.11}$$

and

$$(1+\mu)\|\nabla e\|^2_\Omega + \|e^*\|^2_\Omega \geqslant \int\limits_\Omega |\nabla v - y|^2 dx - \frac{1}{\mu}\sum_{i=1}^m \rho_i^2 + \frac{1}{\mu\lambda_{m+1}}\Big(\sum_{i=1}^m \lambda_i\rho_i^2 - \|\mathbb{R}(y,v)\|^2_\Omega\Big). \tag{5.12}$$

Notice that possible overestimation in (5.11) (and underestimation in (5.12)) is generated by the second and third terms in the right hand sides of these estimates. Assume that $v = v_k$ and $y = y_k$, where $v_k \to u$ in $\overset{\circ}{H}{}^1(\Omega)$, $y_k \to p$ in $L_2(\Omega, \mathbb{R}^d)$ and $\|\mathrm{div} y_k\|_\Omega$ is uniformly bounded as $k \to +\infty$. In this case $\rho_i = \int\limits_\Omega (fw_i - (\mathrm{a}\cdot v_k)w_i - y_k\cdot\nabla w_i)dx \to 0$. The last term can be also made arbitrarily small if m is sufficiently large (see Remark 4.6).

Now, we consider estimates that follow from (4.16). Let $\overline{\Omega} = \bigcup\limits_{k=1}^q \overline{\Omega}_k$, where Ω_k are open Lipschitz subdomains with the diameters $d(\Omega_k)$ such that $\Omega_k \cap \Omega_j = \emptyset$ if $k \neq j$ and

$$\psi_k = \begin{cases} 1 & \text{if } x \in \Omega_k, \\ 0 & \text{if } x \notin \Omega_k, \end{cases}$$

We define $\Psi = \mathrm{span}\{\psi_1, \psi_2, ..., \psi_q\}$ and set $\psi = \sum\limits_{k=1}^q \zeta_k\psi_k$. Then (cf. (2.7))

$$\inf_{\psi\in\Psi} \|w-\psi\|^2_{\mathcal{V}} = \inf_{\substack{\zeta_k\in\mathbb{R},\\ k=1,2,\ldots,q}} \sum_{k=1}^q \|w-\zeta_k\|^2_{\Omega_k} \leqslant \sum_{k=1}^q \frac{d^2(\Omega_k)}{\pi^2}\|\nabla w\|^2_{\Omega_k} \leqslant C^2_{\Psi,\nabla}\|\nabla w\|^2_\Omega, \tag{5.13}$$

where $C_{\Psi,\nabla} = \frac{1}{\pi}\max\limits_k\{d(\Omega_k)\}$.

In this case, (5.9) is replaced by the estimate

$$(1-\mu)\|\nabla e\|^2_\Omega + \|e^*\|^2_\Omega \leqslant \|\nabla v - y\|^2_\Omega + \frac{1}{\mu}\|\nabla u_{\mathrm{R},m}\|^2_\Omega$$
$$+\frac{C^2_{\Psi,\nabla}}{\mu}\Big(\|\mathbb{R}(y,v)\|^2_\Omega - T^{-1}r\cdot r + D^{-1}(b-ST^{-1}r)\cdot(b-ST^{-1}r)\Big), \quad (5.14)$$

where $s_{ki} = \int\limits_{\Omega_k}\Delta w_i dx$,and $b_k = \int\limits_{\Omega_k}\mathbb{R}dx$. The estimate (5.14) will give sharper error bounds than (5.14) if Ω is divided into a large amount of subdomains Ω_k having small diameters, so that the constant in (5.13) is essentially smaller than C_F.

5.3. Estimates for the problem (3.19)–(3.21). Consider the case, where $G(y) = \frac{1}{2}(Ay,y)_U$, and $A : U \to U$ is a bounded linear operator satisfying (2.4).

In this case, $G^*(y^*) = \frac{1}{2}(A^{-1}y^*.y^*)$ and (3.26) reads

$$\frac{1}{2}\|\Lambda e\|^2_A + \frac{1}{2}\|e^*\|^2_{A^{-1}} + \omega\mathcal{D}_R(u,\tau) + \omega\mathcal{D}_R(v,\sigma)$$
$$= \frac{1}{2}\|y^* - A\Lambda v\|^2_{A^{-1}} + \omega\mathcal{D}_R(v,\tau) + \langle\mathbb{R}(y^*,\omega\tau),e\rangle \quad (5.15)$$

Here $p^* = A\Lambda u$, $e^* = y^* - p^*$, and τ is an approximation of σ. To estimate $\langle\mathbb{R}(y^*,\omega\tau),e\rangle$ we use Lemma 2 with $\mathrm{B} = A$ and find that

$$|\langle\,\mathbb{R}(y^*,\omega\tau),e\,\rangle| \leqslant \frac{1}{2\mu}\left(\|\Lambda u_{\mathrm{R},m}\|^2_A + \frac{C^2_\Lambda}{\underline{c}_A}\|\mathbb{R}^\perp_m\|^2_{\mathcal{V}}\right) + \frac{\mu}{2}\|\Lambda e\|^2_A, \quad (5.16)$$

where $\|\mathbb{R}^\perp_m\|_{\mathcal{V}}$ is defined by (4.21) and $u_{\mathrm{R},m}$ solves the finite dimensional problem

$$(A\Lambda u_{\mathrm{R},m}, \Lambda w_m)_U = \langle\mathbb{R}(y^*,\omega\tau),w_m\rangle \qquad \forall w_m \in V_m.$$

From (5.15) and (5.16), we deduce the estimates

$$(1-\mu)\|\Lambda e\|^2_A + \|e^*\|^2_{A^{-1}} + 2\omega\mathcal{D}_R(u,\tau) + 2\omega\mathcal{D}_R(v,\sigma) \leqslant \|y^* - A\Lambda v\|^2_{A^{-1}}$$
$$+ 2\omega\mathcal{D}_R(v,\tau) + \frac{1}{\mu}\|\Lambda u_{\mathrm{R},m}\|^2_A + \frac{C^2_\Lambda}{\mu\underline{c}_A}\|\mathbb{R}^\perp_m\|^2_{\mathcal{V}} \quad (5.17)$$

and

$$(1+\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 + 2\omega\mathcal{D}_R(u,\tau) + 2\omega\mathcal{D}_R(v,\sigma) \geqslant \|y^* - A\Lambda v\|_{A^{-1}}^2$$

$$+ 2\omega\mathcal{D}_R(v,\tau) - \frac{1}{\mu}\|\Lambda u_{\mathrm{R},m}\|_A^2 - \frac{C_\Lambda^2}{\mu\underline{c}_A}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2. \quad (5.18)$$

Right hand sides of (5.17) and (5.18) contain $u_{\mathrm{R},m}$ and directly computable quantities.

If $\omega = 0$ then we arrive at the problem (4.22). In this case,

$$|(\mathbb{R}(y^*), e)_{\mathcal{V}}| \leqslant \sum_{i=1}^m \alpha_i\rho_i + \frac{C_\Lambda^2}{2\mu}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2 + \frac{\mu}{2}\|\Lambda e_m^\perp\|_U^2, \quad (5.19)$$

where α_i are defined bty(4.23). Thus, (5.17) and (5.18) are replaced by

$$(1-\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 \leqslant \|y^* - A\Lambda v\|_{A^{-1}}^2 + \sum_{i=1}^m \alpha_i\rho_i + \frac{C_\Lambda^2}{\mu\underline{c}_A}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2,$$

$$(1+\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 \geqslant \|y^* - A\Lambda v\|_{A^{-1}}^2 + \sum_{i=1}^m \alpha_i\rho_i^2 - \frac{C_\Lambda^2}{\mu\underline{c}_A}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2.$$

Consider the case where Ω is a "simple" domain so that the corresponding eigenfunctions w_i are known. Then (4.35) yields the estimate

$$|\langle\, \mathbb{R}(y^*, \omega\tau), e\,\rangle| \leqslant \frac{1}{\underline{c}_A}\left(\sum_{i=1}^m \rho_i^2 + \frac{1}{\lambda_{m+1}}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2\right)^{1/2}\|\Lambda e\|_A$$

and instead of (5.17) and (5.18) we have two–sided estimates

$$(1-\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 + 2\omega\mathcal{D}_R(u,\tau) + 2\omega\mathcal{D}_R(v,\sigma)$$

$$\leqslant \|y^* - A\Lambda v\|_{A^{-1}}^2 + 2\omega\mathcal{D}_R(v,\tau) + \frac{1}{\mu\underline{c}_A^2}\left(\sum_{i=1}^m \rho_i^2 + \frac{1}{\lambda_{m+1}}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2\right) \quad (5.20)$$

and

$$(1+\mu)\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2 + 2\omega\mathcal{D}_R(u,\tau) + 2\omega\mathcal{D}_R(v,\sigma)$$

$$\geqslant \|y^* - A\Lambda v\|_{A^{-1}}^2 + 2\omega\mathcal{D}_R(v,\tau) - \frac{1}{\mu\underline{c}_A^2}\left(\sum_{i=1}^m \rho_i^2 + \frac{1}{\lambda_{m+1}}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2\right), \quad (5.21)$$

where $\|\mathbb{R}_m^\perp\|_{\mathcal{V}}$ is defined by (4.27).

If V consists of the functions vanishing on the boundary, then we can use arguments of Sec. 4.7 and apply similar estimates with eigenfunctions associated with an extended domain $\widehat{\Omega}$.

§6. Comparison of the estimates

Above considered error bounds have been derived from error identities be estimation of the only one term: $(\mathbb{R}, e)_{\mathcal{V}}$. Therefore, their efficiency depends only on overestimation of this term.

Consider the simples estimate (4.2) first. It is natural to characterise the value of overestimation by the quantity

$$\mathcal{O}_1(e, e^*) := \frac{\frac{C_\Lambda}{\sqrt{\underline{c}_A}}\|\mathbb{R}\|_{\mathcal{V}}\|\Lambda e\|_A - |(\mathbb{R}, e)_{\mathcal{V}}|}{\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2}, \tag{6.1}$$

which relates it to the actual error norm.

Consider the estimate (4.20). For $\mathrm{B} = A$ we have a similar quantity

$$\mathcal{O}_2^m(e, e^*) := \frac{\left(\|\Lambda u_{\mathrm{R},m}\|_A^2 + \frac{C_\Lambda^2}{\underline{c}_A}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}^2\right)^{1/2}\|\Lambda e\|_A - |(\mathbb{R}, e)_{\mathcal{V}}|}{\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2}, \tag{6.2}$$

which involves the solution $u_{\mathrm{R},m}$ of the finite dimensional problem (4.10) and depends on the number m.

Computable error bounds based on the estimate (4.24) also generate overestimation. We characterise it by the quantity

$$\mathcal{O}_3^m(e, e^*) := \frac{\left|\sum\limits_{i=1}^m \alpha_i\rho_i + \frac{C_\Lambda}{\sqrt{\underline{c}_A}}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}\|\Lambda e\|_A - (\mathbb{R}, e)_{\mathcal{V}}\right|}{\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2}. \tag{6.3}$$

Here $\|\Lambda e_m^\perp\|_A^2 = \|\Lambda e\|_A^2 - \sum\limits_{i=1}^m \alpha_i^2$.

$$\alpha_i = (A\Lambda e, \Lambda w_i)_U, \quad \rho_i = (\mathbb{R}, w_i)_{\mathcal{V}} = (\Lambda^* e^*, w_i)_{\mathcal{V}} = (e^*, \Lambda w_i)_U,$$
$$g_i = \Lambda^* A\Lambda w_i, \text{ and } r_i = (\mathbb{R}, g_i)_{\mathcal{V}} = (\Lambda^* e^*, g_i)_{\mathcal{V}}.$$

The estimate (4.35) used in Lemma 3 implies the quantity

$$\mathcal{O}_4^m(e, e^*) := \frac{\left(\sum\limits_{i=1}^m \rho_i^2 + \frac{1}{\lambda_{m+1}}\left(\|\mathbb{R}\|_{\mathcal{V}}^2 - \sum\limits_{i=1}^m \lambda_i\rho_i^2\right)\right)^{1/2}\frac{C_\Lambda}{\sqrt{\underline{c}_A}}\|\Lambda e\|_A - |(\mathbb{R}, e)_{\mathcal{V}}|}{\|\Lambda e\|_A^2 + \|e^*\|_{A^{-1}}^2.} \tag{6.4}$$

Finally, overestimation of the method discussed in Sec. 4.5 for the problem $\Lambda^*\Lambda u + \ell = 0$ is given by the quantity

$$\mathcal{O}_5^m(e,e^*) := \frac{\left|\sum\limits_{i=1}^{M} \rho_i(\zeta_i - \gamma_i) + \frac{1}{\sqrt{\lambda_{m+1}}}\|\mathbb{R}_m^\perp\|_{\mathcal{V}}\,\|\Lambda e\|_U - (\mathbb{R},e)_{\mathcal{V}}\right|}{\|\Lambda e\|^2 + \|e^*\|^2}.$$

where $\mathbb{R}_m^\perp = \mathbb{R} - \sum\limits_{i=1}^{m} \lambda_i\rho_i w_i$ and w_i are the eigenfunctions.

To get a presentation on the difference between the estimates, we compute $\mathcal{O}_i$ $i = 1,2,3,4,5$ for the problem

$$\begin{aligned}&(\phi u')' + f = 0 \qquad \text{in } \Omega = (-1,1),\\ &u(-1) = u(1) = 0,\end{aligned}$$

where $\phi \geqslant \phi_0 > 0$ is a differentiable function. In this case, $p = \phi u'$, $\Lambda^*e^* = -e^{*\prime}$, $\ell = -f$, $C_\Lambda = \frac{2}{\pi}$, and $\underline{c}_A = \phi_0$. Let $w_i = \sin\left(\frac{i\pi}{2}(x+1)\right)$.

Then

$$\|\Lambda e\|_A^2 = \int\limits_{-1}^{1} \phi|e'|^2 dx, \qquad \|e^*\|_{A^{-1}}^2 = \int\limits_{-1}^{1} \phi^{-1}|e^*|^2 dx.$$

and the coefficients are defined by the formulas

$$\alpha_i = \int\limits_{-1}^{1} \phi e' w_i' dx,\ \rho_i = \int\limits_{-1}^{1} e^* w_i' dx,\ g_i = -(\phi w_i')',\ r_i = \int\limits_{-1}^{1} g_i e^{*\prime} dx.$$

Fig. 1 corresponds to the case, where $\phi(x) = 1 + x$ and v is a piecewise affine interpolant of the exact solution u. The functions u and v are depicted in the left part of 1. The right part shows the behaviour of $\mathcal{O}_i$. The value of $\mathcal{O}_1$ is depicted by stars. It does not depend on m and shows maximal level of overestimation. Approximation depicted in Fig. 2 is rather coarse. Nevertheless, the estimates $\mathcal{O}_2$. and $\mathcal{O}_3$ provide good results, which are improving as m grows. The estimates $\mathcal{O}_4$. and $\mathcal{O}_5$ work excellent with very small overestimation. Fig. 3 is related to the case, where approximation is such that the corresponding error has a rather special shape. In spite of this, the estimates provide good results. Certainly, in this case the efficiency of $\mathcal{O}_2 - \mathcal{O}_5$ is different. However for $\mathcal{O}_4$ and $\mathcal{O}_5$ overestimation level is minimal and $\mathcal{O}_3$ also demonstrate sharp estimates when $m > 40$. In general, the results lead to the conclusion that the estimates discussed in Sec. 4 are robust with respect to approximation type and are efficient

Figure 1

Figure 2

for the functions close to the exact solution as well as for coarse approximations. Further verification of them in application to more complicated multydimensional problems is the subject of subsequent publications.

Figure 3

References

1. M. Ainthworth, J. T. Oden, *A posteriori error estimation in finite element analysis*, Wiley, 2000.
2. Д. Е. Апушкинская, С. И. Репин, *Бигармоническая задача с препятствием: гарантированные и вычисляемые оценки ошибок для приближенных решений.* — Ж. вычисл. матем. и матем. физ. **60**, No. 11 (2020), 1881–1897.
3. W. Bangerth, R. Rannacher, *Adaptive finite element methods for differential equations,* Birkhäuser, Berlin, 2003.
4. S. C. Brenner, L. R. Scott, *The mathematical theory of finite element methods*, volume 15. Springer, New York, third edition, 2008.
5. F. Brezzi, M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer Series in Computational Mathematics, 15, New York, 1991.
6. 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**, No. 239 (2002), 945–969.
7. C. Carstensen, *A posteriori error estimate for the mixed finite element method.* — Math. Comp. **66**, No. 218 (1997), 465–476.
8. C. Carstensen, M. Feischl, M. Page, and D. Praetorius, *Axioms of adaptivity.* — Comput. Math. Appl. **67** (2014), 1195–1253.
9. P. Ciarlet, *The finite element method for elliptic problems.* North-Holland, 1987.
10. I. Babuška, T. Strouboulis, *The finite element method and its reliability.* Claderon Press, Oxford, 2001.
11. I. Ekeland, R. Temam, *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.
12. S. G. Mikhlin, *Variational Methods in Mathematical Physics,* Pergamon Press, Oxford, 1964.
13. Л. А. Оганесян, Л. А. Руховец, *О вариационно-разностных схемах для линейных эллиптических уравнений второго порядка в двумерной области с*

кусочно-гладкой границей. — Ж. вычисл. матем. и матем. физ. **8**, No. 1 (1968), 97–114.
14. L. E. Payne, H. F. Weinberger, *An optimal Poincaré inequality for convex domains. — Arch. Rat. Mech. Anal.* **5** (1960), 286–292.
15. S. Repin, *A posteriori error estimation for variational problems with uniformly convex functionals.* — Math. Comput. **69**, No. 230 (2000), 481–500.
16. S. Repin, *Two-sided estimates of deviations from exact solutions of uniformly elliptic equations.* . —Trudi St.Petersburg Math. Soc. **9** (2001), 148–179.
17. S. Repin, *Estimates of deviations from exact solutions initial-boundary value problem for the heat equation.* — Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., 13(2),121–133 (2002).
18. S. Repin, *A posteriori estimates for partial differential equations*, Vol. 4 of Radon Series on Computational and Applied Mathematics. Walter de Gruyter GmbH & Co. KG, Berlin, 2008.
19. S. I. Repin, S. A. Sauter, *Accuracy of Mathematical Models.* Dimension Reduction, Homogenization, and Simplification, volume 33 of EMS Tracts Math. Berlin: European Mathematical Society (EMS), 2020.
20. С. И. Репин, *Тождество для отклонений от точного решения задачи* $\Lambda^* A \Lambda u + \ell = 0$ *и его следствия.* — Ж. вычисл. матем. матем. физ. **61**, No. 12 (2021), 1986–2009.
21. С. И. Репин, *Апостериорные тождества для мер отклонений от точных решений нелинейных краевых задач.* — Ж. вычисл. матем. и матем. физ. **63**, No. 6 (2023), 896–919.
22. S. Repin, *A posteriori error identities for parabolic convection–diffusion problems.* — Zap. Nauchn. Semin. POMI **519**, 205–228.
23. S. I. Repin, *A posteriori error identities and estimates of modelling errors.* — Adv. Appl. Mech. **58** (2024).
24. S. Repin, J. Valdman, *Error identities for variational problems with obstacles.* — ZAMM Z. Angew. Math. Mech. **98**, No. 4 (2018), 635–658.
25. V. Thomée, *Galerkin finite element methods for parabolic problems,* Springer Series in Computational Mathematics, 25, Springer, Berlin,2006.
26. R. Verfürth, *A review of a posteriori error estimation and adaptive mesh–refinement techniques*, Wiley, Teubner, New-York, 1996.
27. O.C. Zienkiewicz, J. Z. Zhu, *A simple error estimator and adaptive procedure for practical engineering analysis.* — Int. J. Num. Meth. Engng. **24** (1987), 337–357.

Санкт-Петербургское отделение
Математического института
им. В. А. Стеклова РАН,
набережная реки Фонтанки 27,
191023, Санкт-Петербург, Россия;
Российский университет дружбы народов
E-mail: repin@@pdmi.ras.ru

Поступило 19 ноября 2024 г.