Tomáš Vejchodský
Fast and guaranteed a posteriori error estimator

In: Jan Chleboun and Petr Přikryl and Karel Segeth (eds.): Programs and Algorithms of Numerical Mathematics, Proceedings of Seminar. Dolní Maxov, June 6-11, 2004. Institute of Mathematics AS CR, Prague, 2004.
pp. 257–272.

Persistent URL: http://dml.cz/dmlcz/702803

FAST AND GUARANTEED A POSTERIORI ERROR ESTIMATOR*

Tomáš Vejchodský

Abstract

The equilibrated residual method and the method of hypercircle are popular methods for a posteriori error estimation for linear elliptic problems. Both these methods are intended to produce guaranteed upper bounds of the energy norm of the error, but the equilibrated residual method is guaranteed only theoretically. The disadvantage of the hypercircle method is its globality, hence slowness. The combination of these two methods leads to local, hence fast, and guaranteed a posteriori error estimator.

1. Introduction

The model problem for our thoughts is a linear elliptic problem on a polygonal domain in $\mathbb{R}^2$ with mixed Dirichlet and Neumann boundary conditions. This problem is discretized by the finite element method. The difference between the exact and the finite element solutions is called the discretization error. Henceforward, if we speak about the error, we have in mind the discretization error.

The goal of this paper is to construct fast and guaranteed a posteriori error estimator. The a posteriori error estimator is a quantity, which is computable from the knowledge of approximate solution and input data, and which bounds or approximate suitable norm of the error. By the word guaranteed we mean that the estimator is really upper bound of the error. By the word fast we mean that the number of arithmetic operations needed for calculation of the estimator is linerly proportional to the number of unknowns.

Note that the finite element solution is the mathematically exact solution of the discrete problem. This solution differs from the numerical solution delivered by the computer, due to round-off errors, quadrature errors and the iteration errors (the errors in the iterative solver of linear algebraic systems). In the current paper, we assume that all arithmetic operations, all integrals and all linear systems are computed exactly. For that reason the computer realization of the proposed methods do not give truly guaranteed error estimators.

In this paper we proposed a posteriori error estimator obtained as a combination of the *equilibrated residual method* and the *method of hypercircle*. The idea to combine these two methods is not new. It has already been suggested in [8], but the a posteriori error estimator proposed there is not completely explicitly computable in 2D and, moreover, the data are assumed to be piecewise constant. The aim of the

*This work was supported by grant No. 201/04/P021 of the Grant Agency of the Czech Republic.

current paper is to show how to construct completely computable a posteriori error estimator for arbitrary data.

The organization of the current paper is following. The linear elliptic model problem is described in Section 2. The equilibrated residual method and the method of hypercircle are briefly introduced in Sections 3 and 4. A posteriori error estimator based on the combination of these two methods is presented in Section 5. Finally, experimental comparison of the mentioned estimators is shown in Section 6.

2. Model problem and notation

Let us consider the following classical formulation of the *linear elliptic problem* in two dimensions

$$\begin{aligned} -\nabla \cdot (\mathcal{A}\nabla \bar{u}) &= f \quad && \text{in } \Omega, \\ \bar{u} &= g_{\mathrm{D}} \quad && \text{on } \Gamma_{\mathrm{D}}, \\ (\mathcal{A}\nabla \bar{u}) \cdot \nu &= g_{\mathrm{N}} \quad && \text{on } \Gamma_{\mathrm{N}}, \end{aligned} \tag{1}$$

where $\Omega \subset \mathbb{R}^2$ is a polygonal domain with Lipschitz boundary $\partial\Omega$, ν is an outer unit normal to $\partial\Omega$, $\overline{\Gamma_{\mathrm{D}}} \cup \overline{\Gamma_{\mathrm{N}}} = \partial\Omega$, $\Gamma_{\mathrm{D}} \cap \Gamma_{\mathrm{N}} = \emptyset$, Γ_{N} has only a finite number of components and the dot symbolizes the Euclidean inner product.

For the purpose of the weak formulation, let us assume that the matrix $\mathcal{A}$ is symmetric and uniformly positive definite with entries in $L^\infty(\Omega)$ and that the function g_{D} is also defined in the interior of Ω such that $g_{\mathrm{D}} \in H^1(\Omega)$. The weak solution $\bar{u} \in H^1(\Omega)$ of problem (1) is given as a sum $\bar{u} = u + g_{\mathrm{D}}$, where $u \in V = \{v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_{\mathrm{D}}\}$ satisfies

$$(\mathcal{A}\nabla u, \nabla v) = (f, v) - (\mathcal{A}\nabla g_{\mathrm{D}}, \nabla v) + \langle g_{\mathrm{N}}, v\rangle \quad \forall v \in V. \tag{2}$$

The right-hand side f and the Neumann boundary condition g_{N} are assumed to be in $L^2(\Omega)$ and $L^2(\Gamma_{\mathrm{N}})$, respectively. The following notation is used

$$\begin{aligned} (\mathcal{A}\nabla u, \nabla v) &= \int_\Omega (\mathcal{A}\nabla u) \cdot \nabla v \,\mathrm{d}x, && u, v \in V, \\ (f, v) &= \int_\Omega f v \,\mathrm{d}x, && f, v \in L^2(\Omega), \\ \langle g_{\mathrm{N}}, v\rangle &= \int_{\Gamma_{\mathrm{N}}} g_{\mathrm{N}} v \,\mathrm{d}s, && g_{\mathrm{N}}, v \in L^2(\Gamma_{\mathrm{N}}). \end{aligned}$$

Note that the L^2-inner product for vector functions from $[L^2(\Omega)]^2$ and for scalar functions from $L^2(\Omega)$ is denoted by the same symbol.

Let us consider an approximate solution $\bar{u}_h = u_h + g_{\mathrm{D}}$ of problem (2), where $\bar{u}_h \in H^1(\Omega)$ and $u_h \in V_h$. The finite element space $V_h \subset V$ is based on a triangulation T_h and consists of continuous and piecewise polynomial functions of degree p. The finite element solution $u_h \in V_h$ is defined by

$$(\mathcal{A}\nabla u_h, \nabla v_h) = (f, v_h) - (\mathcal{A}\nabla g_{\mathrm{D}}, \nabla v_h) + \langle g_{\mathrm{N}}, v_h\rangle \quad \forall v_h \in V_h. \tag{3}$$

The well known *residual equation* follows from (2):

$$(\mathcal{A}\nabla e, \nabla v) = \mathcal{R}(v) \quad \forall v \in V, \tag{4}$$

where $e = u - u_h = \bar{u} - \bar{u}_h$ is the error of the approximate solution and the right-hand side is called *residuum* and is given by

$$\mathcal{R}(v) = (f, v) - (\mathcal{A}\nabla \bar{u}_h, \nabla v) + \langle g_{\mathrm{N}}, v\rangle \quad \forall v \in V.$$

We have the L^2-norm $\|v\|_{0,\Omega}^2 = (v, v)$ and the energy norm $\|v\|^2 = (\mathcal{A}\nabla v, \nabla v)$ for all $v \in V$.

3. The equilibrated residual method

The origin of equilibration residual method goes back to works [8], [4], and [3]. It is not our aim to to describe this method in full detail here. We present only a short overview of the main features of this method. The detailed description of this method can be found in books [1] and [2].

First, let us introduce the local energy norm $\|v\|_K^2 = (\mathcal{A}\nabla v, \nabla v)_K$ on triangle $K \in T_h$ and the local inner products

$$\begin{aligned}
(\mathcal{A}\nabla u, \nabla v)_K &= \int_K (\mathcal{A}\nabla u) \cdot \nabla v \,\mathrm{d}x, \quad && u, v \in H^1(K),\\
(f, v)_K &= \int_K f v \,\mathrm{d}x, && f, v \in L^2(K),\\
\langle g_K, v\rangle_{\partial K} &= \int_{\partial K} g_K v \,\mathrm{d}s, && g_K, v \in L^2(\partial K).
\end{aligned}$$

The error estimator is defined by the following sum

$$\mathcal{E}_{\mathrm{EQ}}^2 = \sum_{K \in T_h} \|\Phi_K\|_K^2 \,, \tag{5}$$

where the functions $\Phi_K \in V(K) = \{v \in H^1(K) : v = 0 \text{ on } \Gamma_{\mathrm{D}}\}$ are the solutions of the local residual problems

$$(\mathcal{A}\nabla \Phi_K, \nabla v)_K = \underbrace{(f, v)_K - (\mathcal{A}\nabla \bar{u}_h, \nabla v)_K + \langle g_K, v\rangle_{\partial K}}_{\mathcal{R}_K^{\mathrm{EQ}}(v)} \quad \forall v \in V(K). \tag{6}$$

The right-hand side of (6) is called local (or equilibrated) residuum and is denoted by $\mathcal{R}_K^{\mathrm{EQ}}(v)$.

The key role in the definition of the equilibrated residuum play the *boundary fluxes* g_K. These boundary fluxes approximate the actual fluxes of the true solution on the element boundaries

$$g_K \approx \nabla u \cdot \nu_K \text{ on } \partial K,$$

where ν_K denotes the outer unit normal to ∂K. The boundary fluxes are constructed as polynomials of degree p on edges which are not on Γ_{N}. These fluxes satisfy the conditions

$$g_K + g_{K^*} = 0 \text{ on } \partial K \cap \partial K^*, \tag{7}$$

where K and K^* are two adjacent elements, and

$$g_K = g_{\mathrm{N}} \text{ on } \partial K \cap \partial \Gamma_{\mathrm{N}}. \tag{8}$$

It is easily seen that conditions (7) and (8) imply equality

$$\mathcal{R}(v) = \sum_{K \in T_h} \mathcal{R}_K^{\mathrm{EQ}}(v|_K) \quad \forall v \in V. \tag{9}$$

We say that $\mathcal{R}_K^{\mathrm{EQ}}(v)$ forms exact splitting of the residuum $\mathcal{R}(v)$. Very important is that the equilibrated fluxes g_K can be determined in such a way that they satisfy the so-called *p-th order equilibration condition*:

$$\mathcal{R}_K^{\mathrm{EQ}}(\theta_K) = (f, \theta_K)_K - (\mathcal{A}\nabla \bar{u}_h, \nabla \theta_K)_K + \langle g_K, \theta_K \rangle_{\partial K} = 0 \tag{10}$$

for all finite element basis functions θ_K on the element K, i.e., for all polynomials of degree p on the element K. Note that the procedure of computing boundary fluxes g_K is not easy and it is not our aim to describe it in this paper. The reader can find this procedure, e.g., in [1].

Let us remark that problem (6) corresponds to the local Neumann problem with boundary conditions given by g_K. This Neumann problem has a solution only if

$$(f, 1)_K + \langle g_K, 1 \rangle_{\partial K} = 0,$$

but this is satisfied thanks to equilibration condition (10). The solution of this local problem is unique up to an additional constant, which is irrelevant, since we are interested only in the gradient of Φ_K, see (5).

Now, let us show that $\mathcal{E}_{\mathrm{EQ}}$ given by (5) is guaranteed upper bound of the energy norm of error. First, we rewrite residual equation (4):

$$(\mathcal{A}\nabla e, \nabla v) = \sum_{K \in T_h} \mathcal{R}_K^{\mathrm{EQ}}(v) = \sum_{K \in T_h} (\mathcal{A}\nabla \Phi_K, \nabla v)_K \quad \forall v \in V.$$

The next step is to employ two times the Cauchy-Schwarz inequality:

$$|(\mathcal{A}\nabla e, \nabla v)| \leq \sum_{K \in T_h} \|\Phi_K\|_K \, \|v\|_K \leq \left(\sum_{K \in T_h} \|\Phi_K\|_K^2 \right)^{1/2} \|v\| \,.$$

This implies

$$\|e\| = \sup_{0 \neq v \in V} \frac{|(\mathcal{A}\nabla e, \nabla v)|}{\|v\|} \leq \left(\sum_{K \in T_h} \|\Phi_K\|_K^2 \right)^{1/2}. \tag{11}$$

Thus, the local solutions Φ_K provide guaranteed a posteriori error estimator. The trouble with the error estimate (11) is that the functions Φ_K as solutions of infinitely dimensional problems (6) are not computable. In practice, we usually use a higher order finite element approximation of Φ_K in (11). Therefore, the practical implementation of the a posteriori error estimator (11) is not guaranteed upper bound.

4. The method of hypercircle

The method of hypercircle is quite old and has many application. Let us mention the fundamental book [11].

Let us derive a guaranteed and computable upper bound employing the method of hypercircle. The standard procedure in the method of hypercircle is the construction of the dual problem in the sense of the calculus of variations, see for example [9], [10], [5], [6], etc. We derive the a posteriori error estimator directly without the definition of the dual problem.

Recall that $H(\mathrm{div},\Omega) \subset [L^2(\Omega)]^2$ denotes the well-known space of functions with divergence in $L^2(\Omega)$. Substituting $v = e = \bar{u} - \bar{u}_h$ into weak formulation (2), we get:

$$-(\mathcal{A}\nabla\bar{u}, \nabla e) = -(f, e) - \langle g_{\mathrm{N}}, e\rangle\,. \tag{12}$$

Introducing a norm $\|\mathbf{q}\|^2_{\mathcal{A}^{-1},\Omega} = (\mathcal{A}^{-1}\mathbf{q}, \mathbf{q})$ for $\mathbf{q} \in H(\mathrm{div},\Omega)$ and employing (12), we can compute

$$\begin{aligned}
\|\mathbf{q} - \mathcal{A}\nabla\bar{u}_h\|^2_{\mathcal{A}^{-1},\Omega} &= \left(\mathcal{A}^{-1}\mathbf{q} - \nabla\bar{u} - \nabla\bar{u}_h + \nabla\bar{u}, \mathbf{q} - \mathcal{A}\nabla\bar{u} - \mathcal{A}\nabla\bar{u}_h + \mathcal{A}\nabla\bar{u}\right)\\
&= \|\mathbf{q} - \mathcal{A}\nabla\bar{u}\|^2_{\mathcal{A}^{-1},\Omega} + 2\,(\mathbf{q} - \mathcal{A}\nabla\bar{u}, \nabla\bar{u} - \nabla\bar{u}_h) + \|\bar{u} - \bar{u}_h\|^2\\
&= \|\mathbf{q} - \mathcal{A}\nabla\bar{u}\|^2_{\mathcal{A}^{-1},\Omega} + 2\,(\mathbf{q}, \nabla e) - 2\,(f, e) - 2\,\langle g_{\mathrm{N}}, e\rangle + \|\bar{u} - \bar{u}_h\|^2
\end{aligned}$$

for arbitrary $\mathbf{q} \in H(\mathrm{div},\Omega)$. Defining the following space

$$Q(f, g_{\mathrm{N}}) = \{\mathbf{q} \in H(\mathrm{div},\Omega) : (\mathbf{q}, \nabla v) = (f, v) + \langle g_{\mathrm{N}}, v\rangle \quad \forall v \in V\}\,,$$

we immediately conclude that

$$\|\mathbf{q} - \mathcal{A}\nabla\bar{u}_h\|^2_{\mathcal{A}^{-1},\Omega} = \|\mathbf{q} - \mathcal{A}\nabla\bar{u}\|^2_{\mathcal{A}^{-1},\Omega} + \|\bar{u} - \bar{u}_h\|^2 \quad \forall \mathbf{q} \in Q(f, g_{\mathrm{N}}).$$

Thus, any $\mathbf{q} \in Q(f, g_{\mathrm{N}})$ gives the guaranteed upper bound

$$\|e\|^2 = \|\bar{u} - \bar{u}_h\|^2 \le \|\mathbf{q} - \mathcal{A}\nabla\bar{u}_h\|^2_{\mathcal{A}^{-1},\Omega}\,. \tag{13}$$

Note that this estimator is exact if $\mathbf{q} = \mathcal{A}\nabla\bar{u}$, but it is unreachable, in general. The question is, how to find a suitable function $\mathbf{q} \in Q(f, g_{\mathrm{N}})$, which would produce a tight upper bound in (13). The crucial ingredient for the answer is the structure of $Q(f, g_{\mathrm{N}})$:

$$Q(f, g_{\mathrm{N}}) = \bar{\mathbf{p}} + \mathbf{curl}\, W, \tag{14}$$

where $\bar{\mathbf{p}} \in Q(f, g_{\mathrm{N}})$ is arbitrary but fixed and

$$W = \left\{ v \in H^1(\Omega) : v = 0 \text{ on } \Gamma_{\mathrm{N}} \right\}.$$

Note that the operator **curl** is defined by $\mathbf{curl} = (\partial/\partial x_2, -\partial/\partial x_1)^\top$ and that

$$\mathbf{curl}\, W = Q(0,0) = \{ \mathbf{q} \in H(\mathrm{div}, \Omega) : (\mathbf{q}, \nabla v) = 0 \quad \forall v \in V \}.$$

This structure of $Q(f, g_{\mathrm{N}})$ is proven in [7] as well as the way how to construct $\bar{\mathbf{p}}$:

$$\bar{\mathbf{p}} = \mathbf{F} + \mathbf{curl}\, w, \tag{15}$$

where

$$\mathbf{F}(x_1, x_2) = \left(- \int_0^{x_1} f(s, x_2)\, \mathrm{d}s, 0 \right)^\top \tag{16}$$

and $w \in H^1(\Omega)$ is an arbitrary function satisfying

$$\mathbf{curl}\, w \cdot \nu = \nabla w \cdot \tau = g_{\mathrm{N}} - \mathbf{F} \cdot \nu \text{ on } \Gamma_{\mathrm{N}}, \tag{17}$$

where $\tau = (-\nu_2, \nu_1)$ is a unit tangent vector to Γ_{N}. We see that the tangent derivative of w is given by $g_{\mathrm{N}} - \mathbf{F} \cdot \nu$. Thus, the values of w on ∂K are given by the primitive function to $g_{\mathrm{N}} - \mathbf{F} \cdot \nu$.

Thanks to (14) we can rewrite estimate (13) as follows

$$\|\bar{u} - \bar{u}_h\|^2 \leq \|\bar{\mathbf{p}} + \mathbf{curl}\, y - \mathcal{A} \nabla \bar{u}_h\|^2_{\mathcal{A}^{-1}, \Omega} \quad \forall y \in W. \tag{18}$$

Now, it is natural to replace W by a finite dimensional subspace $W_h \subset W$ in order to obtain a computable estimate. The easiest way is to construct W_h as a finite element space based on the same triangulation with the same polynomial degree as V_h. The optimal choice $y_h \in W_h$, which minimizes the right-hand side of (18) over W_h satisfies

$$(\mathcal{A}^{-1}\, \mathbf{curl}\, y_h, \mathbf{curl}\, v_h) = (\nabla \bar{u}_h - \mathcal{A}^{-1} \bar{\mathbf{p}}, \mathbf{curl}\, v_h) \quad \forall v_h \in W_h. \tag{19}$$

Note that the idea to minimize the right-hand side of (18) over a finite dimensional space is used, e.g., in [12].

Thus, in order to construct the computable guaranteed upper bound of the energy norm of the error we construct the function $\bar{\mathbf{p}}$ according to (15), (16), and (17). Then, we find solution $y_h \in W_h$ of the finite dimensional problem (19) and compute the norm $\|\bar{\mathbf{p}} + \mathbf{curl}\, y_h - \mathcal{A} \nabla \bar{u}_h\|_{\mathcal{A}^{-1}, \Omega}$, which gives the guaranteed a posteriori error estimator – see (18).

The disadvantage of this procedure is that problem (19) is global and its solution is time-consuming for large problems.

5. The combined method

In order to obtain the locally computable guaranteed upper bound, we combine the local but not guaranteed equilibrated residual method with the guaranteed but not local hypercircle method. The idea is straightforward – compute the boundary fluxes g_K by the equilibrated residual method and then apply the method of hypercircle to local residual problems (6). To do that let us rewrite the equality (6) with $v = \Phi_K$:

$$-(\mathcal{A}\nabla\Phi_K, \nabla\Phi_K)_K = -(f, \Phi_K)_K + (\mathcal{A}\nabla\bar{u}_h, \nabla\Phi_K)_K - \langle g_K, \Phi_K\rangle_{\partial K}. \tag{20}$$

Now, let us compute the following norm for any $\mathbf{q} \in H(\mathrm{div}, K)$:

$$\begin{aligned}\|\mathbf{q}\|^2_{\mathcal{A}^{-1},K} &= \|\mathbf{q} - \mathcal{A}\nabla\Phi_K\|^2_{\mathcal{A}^{-1},K} + 2\,(\mathbf{q} - \mathcal{A}\nabla\Phi_K, \nabla\Phi_K)_K + \|\Phi_K\|^2_K \\ &= 2\,(\mathbf{q}, \nabla\Phi_K)_K - 2(f, \Phi_K)_K + 2(\mathcal{A}\nabla\bar{u}_h, \nabla\Phi_K)_K - 2\langle g_K, \Phi_K\rangle_{\partial K} \\ &\quad + \|\mathbf{q} - \mathcal{A}\nabla\Phi_K\|^2_{\mathcal{A}^{-1},K} + \|\Phi_K\|^2_K\,, \end{aligned} \tag{21}$$

where (20) is utilized. By analogy with the previous section we define the following space

$$\begin{aligned}Q_K(f, g_K, \bar{u}_h) = \{&\mathbf{q} \in H(\mathrm{div}, K) : \\ &(\mathbf{q}, \nabla v)_K = (f, v)_K - (\mathcal{A}\nabla\bar{u}_h, \nabla v)_K + \langle g_K, v\rangle_{\partial K} \quad \forall v \in V(K)\}\,.\end{aligned}$$

The structure of this space is the same as the structure of $Q(f, g_{\mathrm{N}})$ in the previous section. Namely,

$$Q_K(f, g_K, \bar{u}_h) = \bar{\mathbf{p}}_K + \mathbf{curl}\, W(K),$$

where $\bar{\mathbf{p}}_K$ is arbitrary but fixed element of $Q_K(f, g_K, \bar{u}_h)$ and

$$\begin{aligned}W(K) &= \left\{v \in H^1(K) : v = 0 \text{ on } \partial K \setminus \Gamma_{\mathrm{D}}\right\}, \\ \mathbf{curl}\, W(K) &= Q_K(0,0,0) = \{\mathbf{q} \in H(\mathrm{div}, K) : (\mathbf{q}, \nabla v)_K = 0 \quad \forall v \in V(K)\}\,.\end{aligned}$$

From (21) we immediately obtain

$$\|\mathbf{q}\|^2_{\mathcal{A}^{-1},K} = \|\mathbf{q} - \mathcal{A}\nabla\Phi_K\|^2_{\mathcal{A}^{-1},K} + \|\Phi_K\|^2_K \quad \forall \mathbf{q} \in Q_K(f, g_K, \bar{u}_h).$$

Hence,

$$\|\Phi_K\|_K \le \|\mathbf{q}\|_{\mathcal{A}^{-1},K} \quad \forall \mathbf{q} \in Q_K(f, g_K, \bar{u}_h). \tag{22}$$

Finally, using (11), (22), and the structure of the space $Q_K(f, g_K, \bar{u}_h)$, we conclude that

$$\|e\|^2 \le \sum_{K\in T_h} \|\Phi_K\|^2_K \le \sum_{K\in T_h} \|\mathbf{q}\|^2_{\mathcal{A}^{-1},K} = \sum_{K\in T_h} \|\bar{\mathbf{p}}_K + \mathbf{curl}\, y_K\|^2_{\mathcal{A}^{-1},K} \tag{23}$$

holds for all $y_K \in W(K)$ with $\mathbf{q} = \bar{\mathbf{p}}_K + \mathbf{curl}\, y_K$.

In order to evaluate the upper bound (23) we consider a finite dimensional subspace $W_h(K) \subset W(K)$. The subspace $W_h(K)$ may consist, for example, of polynomials of degree $p+1$ on the triangle K. The optimal choice for $y_K \in W_h(K)$, which minimizes the right-hand side of (23) over $W_h(K)$, satisfies

$$\left(\mathcal{A}^{-1}\,\mathbf{curl}\, y_K, \mathbf{curl}\, v\right)_K = -\left(\mathcal{A}^{-1}\bar{\mathbf{p}}_K, \mathbf{curl}\, v\right)_K \quad \forall v \in W_h(K). \tag{24}$$

The only remaining question is how to find the vector $\bar{\mathbf{p}}_K \in Q_K(f, g_K, \bar{u}_h)$ efficiently. The clue is given in the previous section, see (15), (16), and (17):

$$\bar{\mathbf{p}}_K = \mathbf{F} + \mathbf{curl}\, w_K - \mathcal{A}\nabla\bar{u}_h, \tag{25}$$

where $\mathcal{A}\nabla\bar{u}_h$ is known, $\mathbf{F}$ is given by (16) and $w_K \in H^1(K)$ has to satisfy

$$\mathbf{curl}\, w_K \cdot \nu_K = \frac{\partial w_K}{\partial \tau_K} = g_K - \mathbf{F}\cdot\nu_K \quad \text{on } \partial K \setminus \Gamma_{\mathrm{D}}, \tag{26}$$

where the tangent vector is given by $\tau_K = (-\nu_{K,2}, \nu_{K,1})^\top$. Notice that the values of w_K on the boundary ∂K are given by the primitive function to $g_K - \mathbf{F}\cdot\nu_K$. For example, if we consider the triangle K with vertices A, B, C, then the values $w_K(x)$ for $x \in \partial K$ are given by

$$w_K(x) = \begin{cases} w_K(A) + \int_A^x (g_K - \mathbf{F}\cdot\nu_K)\,\mathrm{d}s & \text{for } x \in \overline{AB}, \\ w_K(B) + \int_B^x (g_K - \mathbf{F}\cdot\nu_K)\,\mathrm{d}s & \text{for } x \in \overline{BC}, \\ w_K(C) + \int_C^x (g_K - \mathbf{F}\cdot\nu_K)\,\mathrm{d}s & \text{for } x \in \overline{CA}, \end{cases}$$

where $\overline{AB}$, $\overline{BC}$, and $\overline{CA}$ denote the edges of the triangle and the integration is along these edges. The value $w_K(A)$ is arbitrary and the values $w_K(B)$ and $w_K(C)$ are chosen such that w_K is continuous in the points B and C. Note that the particular value of $w_K(A)$ is irrelevant, because we are interested only in $\mathbf{curl}\, w_K$. The important fact is that the values of function w_K on ∂K are continuous also in the point A: Denoting $w_K(A_-) = \lim\limits_{x\to A, x\in\overline{AB}} w_K(x)$ and $w_K(A_+) = \lim\limits_{x\to A, x\in\overline{CA}} w_K(x)$, we find that

$$w_K(A_+) - w_K(A_-) = \int_{\partial K} \frac{\partial w_K}{\partial \tau_K} = \int_{\partial K} (g_K - \mathbf{F}\cdot\nu_K)\,\mathrm{d}s = \int_{\partial K} g_K\,\mathrm{d}s + \int_K f\,\mathrm{d}x = 0,$$

where the last equality follows from equilibration condition (10) with $\theta_K = 1$.

Thus, the values of w_K are continuous on the boundary of ∂K and it is possible to prolong them into the interior of K such that the prolongation lies in $H^1(K)$. There are infinitely many possibilities how to do this prolongation, but we need the one which leads to the tight upper bound.

We suggest the prolongation illustrated in Figure 1. Consider the triangle K with vertices A, B, C and continuous function ω defined on ∂K, i.e., $\omega \in C^0(\partial K)$. Let us define the function $\tilde{\omega} \in C^0(\overline{K})$ in the following way. First, construct four auxiliary functions:

- The function $\tilde{\ell}$ is a linear function on $\overline{K}$ such that $\tilde{\ell}(A) = \omega(A)$, $\tilde{\ell}(B) = \omega(B)$, and $\tilde{\ell}(C) = \omega(C)$, i.e.,

$$\tilde{\ell} = \omega(A)\varphi_A + \omega(B)\varphi_B + \omega(C)\varphi_C,$$

  where φ_A, φ_B, and φ_C are standard linear finite element basis functions on the triangle K.

- The function $\tilde{z}_3 \in C^0(\overline{K})$ is zero on $\partial K \setminus \overline{AB}$ and is defined by

$$\begin{aligned} &\tilde{z}_3(X) = \omega(X) - \tilde{\ell}(X) && \text{for } X \in \overline{AB}, \\ &\tilde{z}_3(X) = 0 && \text{for } X \in \overline{BC} \cup \overline{CA}, \\ &\tilde{z}_3(X) = \Big(\omega(D_3) - \tilde{\ell}(D_3)\Big) \frac{|XE_3|}{|D_3E_3|} && \text{for } X \in K, \end{aligned} \tag{27}$$

  where $|XE_3|$ denotes the distance between the points X and E_3. The point D_3 is the intersection of the edge $\overline{AB}$ and the line which goes through the point X and is parallel to the edge $\overline{CA}$. The point E_3 is the intersection of this line with the edge $\overline{BC}$ – see Figure 2. Notice that $\tilde{z}_3$ is linear on the line $\overline{D_3E_3}$.

- The functions $\tilde{z}_1 \in C^0(\overline{K})$, which is zero on $\partial K \setminus \overline{BC}$, and $\tilde{z}_2 \in C^0(\overline{K})$, which is zero on $\partial K \setminus \overline{CA}$, are defined by analogy with the definition of $\tilde{z}_3$.

The function $\tilde{\omega} \in C^0(\overline{K})$ is then defined by

$$\tilde{\omega}(X) = \tilde{\ell}(X) + \tilde{z}_1(X) + \tilde{z}_2(X) + \tilde{z}_3(X), \quad X \in \overline{K}. \tag{28}$$

Notice that $\tilde{\omega}(X) = \omega(X)$ on ∂K. This function $\tilde{\omega}$ will be called *the prolongation of ω into interior of K*, or simply *the prolongation of ω*.

The described prolongation has several important properties, which are introduced in the following lemmas. Denote by $P^p(\Theta)$ the space of polynomials of degree p defined on a set Θ.

Lemma 5.1. Consider a triangle K and $\omega \in C^0(\partial K)$. Moreover, let $\omega|_\gamma \in P^p(\gamma)$ for all edges γ of the triangle K and for arbitrary $p \in \mathbb{N}$. Then the prolongation $\tilde{\omega}$ of the function ω into the interior of K given by (28) is a polynomial of degree p in K, i.e., $\tilde{\omega} \in P^p(K)$.

Proof. Let us assume without loss of generality that K is a reference triangle with the vertices $A = [0, 0]$, $B = [1, 0]$, and $C = [0, 1]$. It is enough to prove that $\tilde{z}_1$, $\tilde{z}_2$ and $\tilde{z}_3$ in (28) are polynomials of degree p. Consider, therefore, a function $z_3 \in C^0(\partial K)$, which is zero on $\overline{BC}$ and $\overline{CA}$, and which is a polynomial of degree p on $\overline{AB}$. Note that the functions

$$\varphi_n^{1\mathrm{D}}(x) = x^n(1 - x), \quad n = 1, 2, \ldots, p - 1,$$

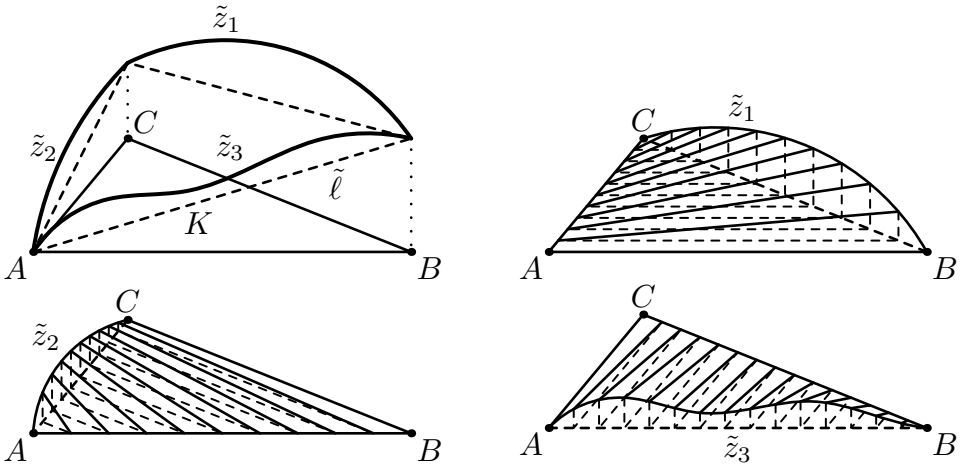

Fig. 1: *The left figure shows the splitting of w_K into the linear part ℓ and to the rest parts z_1, z_2 and z_3 corresponding to edges of triangle K. The right figure illustrates the prolongation of z_1, z_2 and z_3 into the interior of K.*

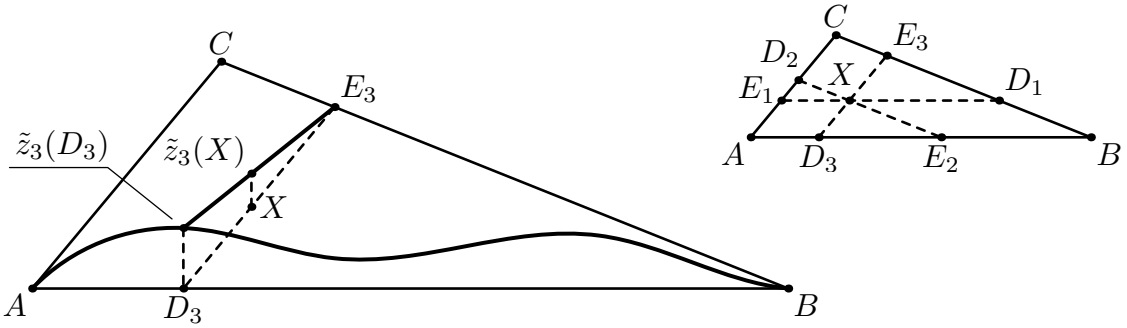

Fig. 2: *The value of z_3 at the interior point x.*

form a basis of the space $P_0^p([0,1])$ of all polynomials on the interval $[0,1]$ with zeros at 0 and 1. Therefore, we can uniquely express the function z_3 on interval $[0,1]$ as a linear combination of these basis functions:

$$z_3 = \sum_{n=1}^{p-1} c_n \varphi_n^{\mathrm{1D}} \quad \text{on } \overline{AB} = [0,1].$$

Now, we can consider the unique prolongation

$$\hat{z}_3 = \sum_{n=1}^{p-1} c_n \varphi_n^{\mathrm{2D}} \quad \text{in } K, \tag{29}$$

where

$$\varphi_n^{\mathrm{2D}}(x_1, x_2) = x_1^n(1 - x_1 - x_2), \quad n = 1, 2, \ldots, p-1,$$

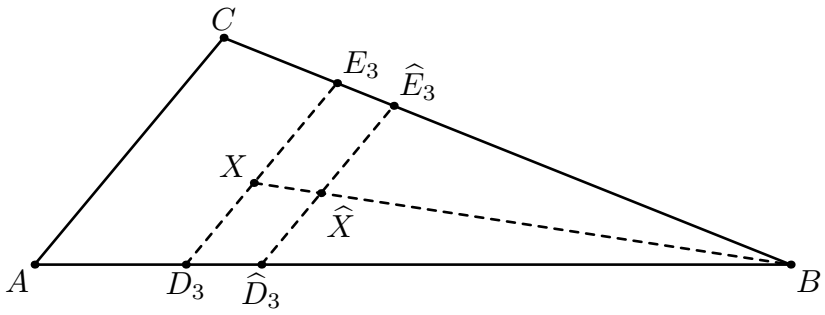

Fig. 3: *The position of points D_3, E_3, $\widehat{D}_3$, $\widehat{E}_3$, X and $\widehat{X}$.*

are the standard finite element basis functions on the reference triangle with zeros on edges $\overline{CA}$ and $\overline{BC}$. Now, consider lines parallel with the edge $\overline{CA}$, i.e., lines described by the equality $x_1 = k$, $k \in \mathbb{R}$. All basis functions $\varphi_n^{2\mathrm{D}}(x_1, x_2)$, $n = 1, 2, \ldots, p-1$, are linear on these lines:

$$\varphi_n^{2\mathrm{D}}(k, x_2) = k^n(1 - k - x_2), \quad n = 1, 2, \ldots, p-1.$$

Therefore, also the function $\hat{z}_3$ is linear on the lines parallel with edge $\overline{CA}$, which is exactly the way, how the prolongation $\tilde{z}_3$ described above is constructed. Hence, the prolongations $\hat{z}_3$ given by (29) and $\tilde{z}_3$ given by (27) of the function z_3 coincide and we conclude that $\tilde{z}_3$ is a polynomial of degree p in K. Analogical reasoning can be done for z_1 and z_2, too. □

Lemma 5.2. Consider a triangle K with vertices A, B, C, function $\omega \in C^0(\partial K)$ and its prolongation $\tilde{\omega} \in C^0(K)$ defined by (28). If there exists a finite tangent derivative $\partial\omega/\partial\tau_K$ on all edges of the triangle K then the derivatives of the function $\tilde{z}_3$, defined by (27), in the directions $\overrightarrow{ED}$ and $\overrightarrow{XB}$ are given at any interior point $X = (x_1, x_2) \in K$ by

$$\frac{\partial \tilde{z}_3(X)}{\partial\, \overrightarrow{E_3D_3}} = \frac{\tilde{z}_3(D_3)}{|D_3E_3|},$$
$$\frac{\partial \tilde{z}_3(X)}{\partial\, \overrightarrow{XB}} = \frac{\partial \tilde{z}_3(D_3)}{\partial\, \overrightarrow{AB}} \frac{|AB|}{|XB|} \frac{|XE_3|}{|D_3E_3|}\alpha = \frac{\partial \tilde{z}_3(D_3)}{\partial \tau_K} \frac{|AB|}{|XB|} \frac{|XE_3|}{|D_3E_3|}\alpha,$$

where

$$\alpha = \frac{(B_1 - x_1)(A_1 - C_1) - (B_2 - x_2)(A_2 - C_2)}{(B_1 - A_1)(A_1 - C_1) - (B_2 - A_2)(A_2 - C_2)}.$$

Proof. Figure 3 illustrates the used notations. The derivative in the directions $\overrightarrow{E_3D_3}$ is easy to calculate. The derivative in the directions $\overrightarrow{XB}$ is given by

$$\lim_{r\to 0}\frac{\tilde{z}_3(X+r(B-X))-\tilde{z}_3(X)}{r\,|BX|}=\lim_{r\to 0}\frac{\tilde{z}_3(\widehat{X})-\tilde{z}_3(X)}{r\,|BX|}=\lim_{r\to 0}\frac{\tilde{z}_3(\widehat{D}_3)-z_3(D_3)}{r\,|BX|}\frac{|XE_3|}{|D_3E_3|},$$

where $\widehat{X}=X+r(B-X)$,

$$\tilde{z}_3(X)=\tilde{z}_3(D_3)\frac{|XE_3|}{|D_3E_3|},\qquad \tilde{z}_3(\widehat{X})=\tilde{z}_3(\widehat{D}_3)\frac{\left|\widehat{X}\widehat{E}_3\right|}{\left|\widehat{D}_3\widehat{E}_3\right|},$$

and

$$\frac{|XE_3|}{|D_3E_3|}=\frac{\left|\widehat{X}\widehat{E}_3\right|}{\left|\widehat{D}_3\widehat{E}_3\right|},$$

which holds thanks to the similarity of appropriate triangles. The rest of the proof is an exercise in the analytical geometry. □

Note that the derivatives of $\tilde{z}_1$ and $\tilde{z}_2$ can be evaluated by analogy with Lemma 5.2.

Let us summarize the way how to compute the guaranteed and locally computable upper bound. First, compute boundary fluxes g_K using residual equilibration method. Then, construct vector $\bar{\mathbf{p}}_K$ for all triangles K in T_h by (25). The values of w_K on ∂K are given by the primitive function to $g_K-\mathbf{F}\cdot\nu$. These values are prolonged into the interior of K according to (28) and this prolongation is denoted also by w_K. Important is that the values of $\mathbf{curl}\,w_K$ are easily computable from values of w_K on ∂K and from $\partial w_K/\partial\tau_K=g_K-\mathbf{F}\cdot\nu$ on ∂K thanks to Lemma 5.2. The next step is to find a solution $y_K\in W_h(K)$ of finite dimensional local problem (24). The final step is to evaluate estimate (23).

Thanks to the properties of prolongation (28) we can prove the following lemma.

Lemma 5.3. Let the finite element solution $u_h\in V_h$ be exact, i.e., $u_h=u$ and let the matrix $\mathcal{A}$ be constant. If the vector $\bar{\mathbf{p}}_K\in Q_K(f,g_K,\bar{u}_h)$ is constructed as described above then the error estimator (23) is exact, i.e., $\bar{\mathbf{p}}_K+\mathbf{curl}\,y_K=0$.

Proof. Consider a triangle $K\in T_h$. The right-hand side f has to be a polynomial of degree at most $p-2$ on K, since $u_h=u$ is a polynomial of degree at most p and $\mathcal{A}$ is constant. Therefore, $\mathbf{F}$ is a polynomial of degree at most $p-1$. The boundary fluxes g_K are constructed as polynomials of degree p on edges of K and they coincide with the actual fluxes of the exact solution:

$$g_K=\nabla\bar{u}\cdot\nu_K \text{ on } \partial K.$$

By the way, this fact implies that $\Phi_K=0$, see (6). Moreover, it implies that

$$Q_K(f,g_K,\bar{u}_h)=\{\mathbf{q}\in H(\mathrm{div},K):(\mathbf{q},\nabla v)_K=0\quad\forall v\in V(K)\}=\mathbf{curl}\,W(K).$$

Thus, the function w_K is a polynomial of degree at most $p+1$ on each edge of K and Lemma 5.1 implies that the prolongation of these values into the interior of K is also a polynomial of degree at most $p+1$.

Hence, $\bar{\mathbf{p}}_K$ is a polynomial of degree at most p and that $\bar{\mathbf{p}}_K \in \mathbf{curl}\, W(K)$. Consequently, $\bar{\mathbf{p}}_K \in \mathbf{curl}\, W_h(K)$ and projection (24) gives $\mathbf{curl}\, y_k = -\bar{\mathbf{p}}_K$. □

6. Numerical experiments

The practical behavior of the combined method is compared with the equilibrated residual method and with the method of hypercircle. The *quadratic finite element method* is chosen. In case of equilibrated residual method, the local residual problem (6) is solved on the space $V_h(K)$ of *cubic* polynomials on triangle K. In case of the combined method, local problem (24) is solved on the space $W_h(K)$ of *cubic* polynomials, too. The difference between local problems for equilibrated residual and combined methods is in boundary conditions and in the number of degrees of freedom. For example, if we consider the interior element K, then $\dim V_h(K) = 10$ and $\dim W_h(K) = 1$. Thus, the combined method performs faster in the solution of local problems. Anyway, the computation of boundary fluxes g_K have to be done for both methods and it takes considerable amount of time.

The global problem (19) in classical method of hypercircle is solved by the *quadratic finite elements*. Note that the spaces V_h and W_h differ due to different boundary conditions for problems (3) and (19). The function w, see (15), is constructed in the similar way as function w_K in Section 5.

Example 6.1. Consider problem (1) with the following data: $\Omega = [-1,1]^2$, $\Gamma_{\mathrm{D}} = \partial\Omega$, $\Gamma_{\mathrm{N}} = \emptyset$, $\mathcal{A}$ is the identity matrix, $f(x_1,x_2) = 2(2 - x_1^2 - x_2^2)$, $g_{\mathrm{D}} = 0$. The exact solution corresponding to this problem is $u(x_1,x_2) = (x_1^2-1)(x_2^2-1)$. Table 1 shows the effectivity indices $I_{\mathrm{eff}} = \mathcal{E}/\,|||e|||$, where the estimator $\mathcal{E}$ is obtained by the equilibrated residual method, by the method of hypercircle and by the combined method, respectively. The results are for the regular meshes. First five of these meshes is illustrated in Figure 4.

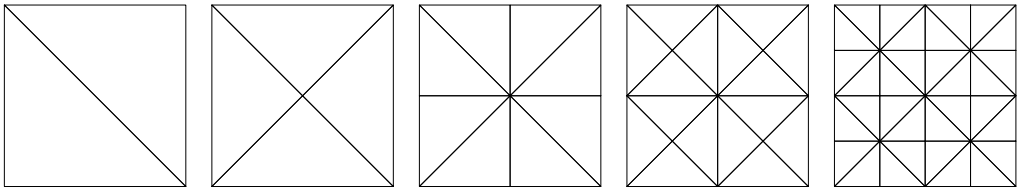

Fig. 4: *Regular triangulations of the domain* $\Omega = [-1,1]^2$ *with 2, 4, 8, 16, and 32 triangles.*

Example 6.2. Let us consider the following data: $\Omega = [0,1]^2$, $\Gamma_{\mathrm{D}} = \{(x_1,x_2) \in \partial\Omega : x_1 = 0 \text{ or } x_2 = 0\}$, $\Gamma_{\mathrm{N}} = \partial\Omega\backslash\Gamma_{\mathrm{D}}$, $\mathcal{A}$ equals to the identity matrix, $g_{\mathrm{D}} = 0$, $c = 0$, $\alpha = 0$, g_{N} and f are such that the exact solution is $u(x_1,x_2) = \sin(17x_1x_2)\mathrm{e}^{x_1+x_2}$.

N_{tri}	equilibrated residua	method of hypercircle	combined method
2	1.43	1.11	1.06
4	1.23	1.25	1.01
8	1.34	1.20	1.00
16	1.30	1.30	1.16
32	1.39	1.23	1.29
64	1.32	1.32	1.27
128	1.41	1.25	1.52
256	1.33	1.34	1.33
512	1.41	1.25	1.64
1024	1.33	1.34	1.36
2048	1.41	1.26	1.71
4096	1.33	1.34	1.38
8192	1.41	1.26	1.74
16384	1.33	1.34	1.38
32768	1.41	1.26	1.75

Tab. 1: *Effectivity indices for Example 6.1 computed by the equilibrated residual method, by the method of hypercircle and by the combined method on regular triangulations of the domain with N_{tri} triangles.*

Table 2 shows the comparison of the effectivity indices of the discussed error estimators on the same regular meshes, see Figure 4. Notice that the equilibrated residual method really underestimates the error for first four meshes.

7. Conclusions

The equilibrated residual method and the classical method of hypercircle have been discussed in this paper. It was shown how to obtain a posteriori upper bounds of energy norm of error by both of these methods. However, the equilibrated residual method gives the upper bound which is not computable, since it involves solutions of infinitely dimensional local problems. In practice, approximations of these local problems are used and, therefore, the computable version is not guaranteed to be an upper bound.

On the other hand, the method of hypercircle produces the guaranteed and computable upper bound, but its evaluation needs the solution of a global problem. Thus, computation of the upper bound by the method of hypercircle is not fast.

The combination of the equilibrated residual method and the method of hypercircle has been suggested in this paper. This combined method provides the guaranteed and locally computable upper bound. Numerical experiments indicate that the combined method performs very well. It is efficient and fast, in the sence that the computational time is linearly proportional to the size of the problem. Of course, the

N_{tri}	equilibrated residua	hypercircle method	combined method
2	0.37	1.93	1.86
4	0.29	1.70	1.71
8	0.49	1.65	1.69
16	0.94	1.51	1.19
32	1.11	1.56	1.37
64	1.06	1.51	1.29
128	1.14	1.68	1.37
256	1.17	1.48	1.28
512	1.30	1.52	1.58
1024	1.22	1.49	1.35
2048	1.34	1.50	1.67
4096	1.24	1.50	1.38
8192	1.35	1.49	1.70
16384	1.25	1.51	1.39
32768	1.35	1.49	1.71

Tab. 2: *Effectivity indices for Example 6.2 computed by the equilibrated residual method, by the method of hypercircle and by the combined method on regular triangulations of the domain with N_{tri} triangles.*

combined method is slower than the equilibrated residual method, since it requires additional work with function w_K, see Figures 1 and 2, but it gives guaranteed upper bound. The efficiency of the combined method, measured by the effectivity index, is comparable with the equilibrated residual method as well as with the method of hypercircle, which both provide tight bounds of the error.

References

[1] M. Ainsworth, J.T. Oden: *A posteriori error estimation in finite element analysis*. John Wiley & Sons, New York, 2000.

[2] I. Babuška, T. Strouboulis: *The finite element method and its reliability*. Oxford University Press, New York, 2001.

[3] R.E. Bank, A. Weiser: *Some a posteriori error estimators for elliptic partial differential equations*. Math. Comp., **44**, 1985, 283–301.

[4] D.W. Kelly: *The self-equilibration of residuals and complementary a posteriori error estimates in the finite element method*. Internat. J. Numer. Methods Engrg., **20**, 1984, 1491–1506.

[5] J. Haslinger, I. Hlaváček: *Convergence of a finite element method based on the dual variational formulation*. Appl. Math., **21**, 1976, 43–65.

[6] I. Hlaváček: *Dual finite element analysis of axisymmetric elliptic problems with an absolute term*. Appl. Math., **36**, 1991, 392–406.

[7] M. Křížek: *Conforming equilibrium finite element methods for some elliptic plane problems*. RAIRO Anal. Numér., **17**, 1983, 35–65.

[8] P. Ladeveze, D. Leguillon: *Error estimate procedure in the finite element method and applications*. SIAM J. Numer. Anal., **20**, 1983, 485–509.

[9] S.I. Repin: *A unified approach to a posteriori error estimation based on duality error majorants*. Math. Comput. Simulation, **50**, 1999, 305–321.

[10] S.I. Repin: *A posteriori error estimation for variational problems with uniformly convex functionals*. Math. Comp., **69**, 1999, 481–500.

[11] J.L. Synge: *The hypercircle in mathematical physics: a method for the approximate solution of boundary value problems*. Cambridge University Press, New York, 1957.

[12] J. Weisz: *On a method for a posteriori error estimation of approximate solutions to parabolic problems*. Comment. Math. Univ. Carolin., **35**, 1994, 735–740.