

Tomáš Vejchodský

On the quality of local flux reconstructions for guaranteed error bounds

In: Jan Brandts and Sergej Korotov and Michal Křížek and Karel Segeth and Jakub Šístek and Tomáš Vejchodský (eds.): Application of Mathematics 2015, In honor of the birthday anniversaries of Ivo Babuška (90), Milan Práger (85), and Emil Vitásek (85), Proceedings. Prague, November 18-21, 2015. Institute of Mathematics CAS, Prague, 2015. pp. 242–255.

Persistent URL: <http://dml.cz/dmlcz/702981>

Terms of use:

© Institute of Mathematics CAS, 2015

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library*
<http://dml.cz>

ON THE QUALITY OF LOCAL FLUX RECONSTRUCTIONS FOR GUARANTEED ERROR BOUNDS

Tomáš Vejchodský

Institute of Mathematics, Czech Academy of Sciences
Žitná 25, Praha 1, Czech Republic
vejchod@math.cas.cz

Abstract: In this contribution we consider elliptic problems of a reaction-diffusion type discretized by the finite element method and study the quality of guaranteed upper bounds of the error. In particular, we concentrate on complementary error bounds whose values are determined by suitable flux reconstructions. We present numerical experiments comparing the performance of the local flux reconstruction of Ainsworth and Vejchodský [2] and the reconstruction of Braess and Schöberl [5]. We evaluate the efficiency of these flux reconstructions by their comparison with the optimal flux reconstruction computed as a global minimization problem.

Keywords: a posteriori error estimates, complementarity, index of effectivity, elliptic problem, reaction-diffusion, singular perturbation

MSC: 65N15, 65N30

1. Introduction

The popularity of the complementary error bounds has grown during recent years due to their favourable properties. They provide guaranteed upper bounds on the error, they are locally efficient, robust, and there are fast algorithms for their computation. The main idea of these error bounds goes back the ‘method of hypercircle’ [15, 19] and it was developed in [3, 8, 9, 10, 12, 23] and other papers. During recent years this idea attracted a lot of attention [13, 16, 20, 21] and it was used even for partial differential eigenvalue problems [18]. Error bounds of this type for reaction-diffusion problems were recently presented in [6, 11, 17, 22, 24] and elsewhere.

However, the papers [1, 2] are the only one (to our knowledge) where a locally computable and robust upper bound on the error of the finite element solution is presented. The main goal of this contribution is to compare the accuracy of this error bound with the bound proposed in [5] for the Poisson problem, see also [7] and [4, Algorithm 9.3].

We consider the following linear elliptic problem of the reaction-diffusion type with mixed boundary conditions:

$$-\Delta u + \kappa^2 u = f \quad \text{in } \Omega; \quad u = 0 \quad \text{on } \Gamma_D; \quad \partial u / \partial \mathbf{n} = g_N \quad \text{on } \Gamma_N. \quad (1)$$

Here, $\Omega \subset \mathbb{R}^2$ is a domain, \mathbf{n} stands for the unit outward-facing normal vector to the boundary $\partial\Omega$, the portions Γ_D and Γ_N of the boundary $\partial\Omega$ are open, disjoint, and satisfy $\overline{\Gamma_D} \cup \overline{\Gamma_N} = \partial\Omega$. We assume the reaction coefficient $\kappa \geq 0$ to be piecewise constant. In order to guarantee unique solvability of (1), we assume that $\kappa > 0$ in a subdomain of Ω of a positive measure or that Γ_D has a positive measure.

In order to discretize this problem by the standard lowest-order finite elements, we approximate the domain Ω by a polygon Ω_h , Γ_D by $\Gamma_{D,h} \subset \partial\Omega_h$, and Γ_N by $\Gamma_{N,h} \subset \partial\Omega_h$. The weak solution in the domain Ω_h is defined as $\tilde{u} \in V = \{v \in H^1(\Omega_h) : v = 0 \text{ on } \Gamma_{D,h}\}$ such that

$$\mathcal{B}(\tilde{u}, v) = \mathcal{F}(v) \quad \forall v \in V, \quad (2)$$

where

$$\mathcal{B}(\tilde{u}, v) = \int_{\Omega_h} (\nabla \tilde{u} \cdot \nabla v + \kappa^2 \tilde{u} v) \, d\mathbf{x} \quad \text{and} \quad \mathcal{F}(v) = \int_{\Omega_h} f v \, d\mathbf{x} + \int_{\Gamma_{N,h}} g_N v \, ds$$

for $\tilde{u}, v \in H^1(\Omega_h)$. We remark that $H^1(\Omega_h)$ stands for the usual Sobolev space $W^{1,2}(\Omega_h)$.

The approximation of the general domain Ω by a polygon Ω_h introduces a boundary approximation error $u - \tilde{u}$. In this paper we strictly distinguish between the solution in Ω and the solution in Ω_h in order to emphasize the fact that the error estimators discussed below do not include the boundary approximation error. Moreover, the numerical examples in Section 5 are posed in a circular disc and, therefore, there is a nonzero boundary approximation error. Such examples enable us to discuss the relative size of the boundary approximation error with respect to the other components of the total error estimated by the computed error bounds.

We discretize problem (2) by the lowest-order finite element method. Therefore, we consider a triangulation \mathcal{T}_h of Ω_h consisting of triangular elements. The union of all triangles in \mathcal{T}_h is $\overline{\Omega_h}$, the interiors of triangles in \mathcal{T}_h are pairwise disjoint, and every edge of each triangle lies either on $\partial\Omega_h$ or it is completely shared by exactly two neighbouring triangles. The discretization parameter is defined as $h = \max_{K \in \mathcal{T}_h} h_K$, where $h_K = \text{diam } K$. We also assume that the triangulation \mathcal{T}_h is compatible with the piecewise constant coefficient κ and denote the value of κ on an element $K \in \mathcal{T}_h$ by κ_K . Using this triangulation, we define the usual finite element space

$$V_h = \{u_h \in V : u_h|_K \in P^1(K) \, \forall K \in \mathcal{T}_h\},$$

where $P^1(K)$ stands for the space of linear functions on the triangle $K \in \mathcal{T}_h$. Finally, the finite element formulation of problem (2) reads: find $u_h \in V_h$ such that

$$\mathcal{B}(u_h, v_h) = \mathcal{F}(v_h) \quad \forall v_h \in V_h. \quad (3)$$

Let us note that problem (1) can be diffusion dominated or singularly perturbed depending on the size of the reaction coefficient κ . The behaviour of the finite element method depends on the size of the discretization parameter h with respect to κ . If κh is small then possible boundary layers, which may occur if κ is large, are well resolved and the finite element solution is accurate. However, if κh is large, then boundary layers are not well captured by the mesh, the finite element solution exhibits spurious oscillations and its error is relatively large. This error behaviour has to be reflected by the error bounds. Therefore, we often distinguish the cases of small and large κh and observe differences in the accuracy.

2. Complementary error bounds

In this section we present two types of complementary error bounds. These bounds are similar, but they slightly differ in definition, assumptions, and applicability. Surprisingly, they differ considerably in performance. One bound provides accurate results for problems with large κh , while the other one for small κh .

First, let us introduce certain notation. Let $\|v\|^2 = \mathcal{B}(v, v)$ be the energy norm and $\|v\|_K$ be the $L^2(K)$ norm of v . Let $\Pi_K f \in P^1(K)$ be $L^2(K)$ -orthogonal projection of f onto $P^1(K)$, $K \in \mathcal{T}_h$. Similarly, if γ is an edge of a triangle $K \in \mathcal{T}_h$, then $\Pi_\gamma g_N$ is the $L^2(\gamma)$ -orthogonal projection of $g_N \in L^2(\gamma)$ onto $P^1(\gamma)$. We also define oscillation terms

$$\text{osc}_K(f) = \min \left\{ \frac{h_K}{\pi}, \frac{1}{\kappa_K} \right\} \|f - \Pi_K f\|_K, \quad \text{osc}_\gamma(g_N) = \min\{C_T, \bar{C}_T\} \|g_N - \Pi_\gamma g_N\|_\gamma,$$

where $K \in \mathcal{T}_h$ and $\gamma \subset \Gamma_{N,h} \cap \partial K$ is an edge. Constants C_T and \bar{C}_T are defined in [2] as

$$C_T^2 = \frac{|\gamma|}{d|K|} \frac{1}{\kappa_K} \sqrt{(2h_K)^2 + (d/\kappa_K)^2},$$

$$\bar{C}_T^2 = \frac{|\gamma|}{d|K|} \min\{h_K/\pi, \kappa_K^{-1}\} (2h_K + d \min\{h_K/\pi, \kappa_K^{-1}\}),$$

where $d = 2$ is the dimension.

To handle the Neumann boundary conditions, we seek the flux reconstruction in

$$\mathbf{W} = \{\boldsymbol{\tau} \in \mathbf{H}(\text{div}, \Omega_h) : \boldsymbol{\tau} \cdot \mathbf{n} = \Pi_\gamma g_N \text{ on all edges } \gamma \subset \Gamma_{N,h} \cap \partial K \text{ of all } K \in \mathcal{T}_h\}.$$

Having a flux reconstruction $\boldsymbol{\tau} \in \mathbf{W}$, we can consider the error bound $\eta(\boldsymbol{\tau})$ in the general form

$$\eta^2(\boldsymbol{\tau}) = \sum_{K \in \mathcal{T}_h} \left[\eta_K(\boldsymbol{\tau}) + \text{osc}_K(f) + \sum_{\gamma \subset \Gamma_{N,h} \cap \partial K} \text{osc}_\gamma(g_N) \right]^2, \quad (4)$$

where $\eta_K(\boldsymbol{\tau})$ is an error indicator computed from the values of $\boldsymbol{\tau}$ restricted to K only. We introduce two error indicators. If $\boldsymbol{\tau}$ on an element K satisfies the equilibration condition

$$\int_K (\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}|_K) \, d\mathbf{x} = 0, \quad (5)$$

then we set

$$\eta_K^a(\boldsymbol{\tau}) = \|\boldsymbol{\tau} - \nabla u_h\|_K + \frac{h_K}{\pi} \|\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}\|_K, \quad (6)$$

otherwise $\eta_K^a(\boldsymbol{\tau})$ is undefined. If $\kappa_K > 0$, then we put

$$\eta_K^b(\boldsymbol{\tau}) = \left(\|\boldsymbol{\tau} - \nabla u_h\|_K^2 + \kappa_K^{-2} \|\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}\|_K^2 \right)^{1/2}, \quad (7)$$

otherwise $\eta_K^b(\boldsymbol{\tau})$ is undefined. The following theorem proves that both these error indicators provide guaranteed upper bounds on the energy norm of the error.

Theorem 1. *Let $\tilde{u} \in V$ be the weak solution (2). Let both $u_h \in V$ and $\boldsymbol{\tau} \in \mathbf{W}$ be arbitrary. Then*

$$\|\tilde{u} - u_h\| \leq \eta^{\text{ab}}(\boldsymbol{\tau}), \quad (8)$$

where $\eta^{\text{ab}}(\boldsymbol{\tau})$ is given by (4) with

$$\eta_K(\boldsymbol{\tau}) = \begin{cases} \min\{\eta_K^a(\boldsymbol{\tau}), \eta_K^b(\boldsymbol{\tau})\} & \text{if (5) holds in } K \in \mathcal{T}_h \text{ and } \kappa_K > 0, \\ \eta_K^a(\boldsymbol{\tau}) & \text{if (5) holds in } K \in \mathcal{T}_h \text{ and } \kappa_K = 0, \\ \eta_K^b(\boldsymbol{\tau}) & \text{if (5) does not hold in } K \in \mathcal{T}_h \text{ and } \kappa_K > 0, \end{cases}$$

Proof. Proofs of different variants of this theorem can be found in many places in the literature. The main idea traces back to the method of hypercircle and [19]. For the reader's convenience we briefly present the main steps of the proof and refer to [2] for details.

Let $v \in V$ be arbitrary. Using the weak formulation (2) for \tilde{u} , splitting the integrals in definitions of \mathcal{B} and \mathcal{F} into sums over all elements in \mathcal{T}_h , and applying the divergence theorem for $\boldsymbol{\tau} \in \mathbf{W}$, we obtain the identity

$$\begin{aligned} \mathcal{B}(\tilde{u} - u_h, v) = \sum_{K \in \mathcal{T}_h} \left[\int_K (\boldsymbol{\tau} - \nabla u_h) \cdot \nabla v \, d\mathbf{x} + \int_K (\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}) v \, d\mathbf{x} \right. \\ \left. + \int_K (f - \Pi_K f) v \, d\mathbf{x} + \sum_{\gamma \subset \Gamma_{N,h} \cap \partial K} \int_{\gamma} (g_N - \Pi_{\gamma} g_N) v \, d\mathbf{s} \right]. \quad (9) \end{aligned}$$

For brevity, let us denote $\mathbf{g} = \boldsymbol{\tau} - \nabla u_h$ and $r_K = \Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}$. If the equilibration condition (5) is satisfied in K then we obtain

$$\int_K r_K v \, d\mathbf{x} \leq \int_K r_K (v - \bar{v}_K) \, d\mathbf{x} \leq \|r_K\|_K \|v - \bar{v}_K\|_K \leq \frac{h_K}{\pi} \|r_K\|_K \|\nabla v\|_K,$$

where $\bar{v}_K = |K|^{-1} \int_K v \, d\mathbf{x}$ and we use the Poincaré inequality [14]. Using this estimate, we easily bound the first two integrals on the right-hand side of (9) as

$$\int_K \mathbf{g} \cdot \nabla v \, d\mathbf{x} + \int_K r_K v \, d\mathbf{x} \leq \eta_K^a(\boldsymbol{\tau}) \|\nabla v\|_K \leq \eta_K^a(\boldsymbol{\tau}) \|v\|_K, \quad (10)$$

where $\|v\|_K^2 = \|\nabla v\|_K^2 + \kappa_K^2 \|v\|_K^2$ stands for the local energy norm. Alternatively, if $\kappa_K > 0$, we can bound these two integrals as

$$\int_K \mathbf{g} \cdot \nabla v \, d\mathbf{x} + \int_K r_K v \, d\mathbf{x} \leq \|\mathbf{g}\|_K \|\nabla v\|_K + \|r_K\|_K \|v\|_K \leq \eta_K^b(\boldsymbol{\tau}) \|v\|_K. \quad (11)$$

To finish the proof we use (10) and (11) in (9), estimate the last two integrals on the right-hand side of (9) by the corresponding oscillation terms, see [2], substitute $v = \tilde{u} - u_h$, and apply the Cauchy-Schwarz inequality. \square

Note that neither η_K^a nor η_K^b provide an error bound if (5) does not hold and $\kappa_K = 0$. Further note that the error indicators η_K^a and η_K^b given in (6) and (7) coincide if $\boldsymbol{\tau} \in \mathbf{W}$ is chosen in such a way that $\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau} = 0$ for all $K \in \mathcal{T}_h$. In this case $\eta_K^a(\boldsymbol{\tau}) = \eta_K^b(\boldsymbol{\tau}) = \|\boldsymbol{\tau} - \nabla u_h\|_K$ provides the upper bound (8) even for $\kappa_K = 0$, see [2].

However, in general the indicators (6) and (7) differ. Considering the optimal flux reconstruction, the indicator η_K^a typically yields smaller values than η_K^b for small $\kappa_K h_K$ including $\kappa_K = 0$. However, if $\kappa_K h_K$ is large then η_K^b provides tight and robust upper bound and η_K^a overestimates the error unacceptably. Moreover, formulas (6) and (7) for η_K^a and η_K^b have different structures, which can be unpleasant from the practical point of view. Therefore, we unify both these indicators into a single one, which is comparatively accurate as $\min\{\eta_K^a(\boldsymbol{\tau}), \eta_K^b(\boldsymbol{\tau})\}$, but always (slightly) greater or equal.

Lemma 2. *Let $K \in \mathcal{T}_h$ and let η_K^a and η_K^b be given by (6) and (7). Further, let $\boldsymbol{\tau} \in \mathbf{W}$ and let $\boldsymbol{\tau}|_K$ satisfy the equilibration condition (5). Finally, let $\kappa_K > 0$. Then*

$$\min\{\eta_K^a(\boldsymbol{\tau}), \eta_K^b(\boldsymbol{\tau})\} \leq \eta_K^c(\boldsymbol{\tau}), \quad (12)$$

where

$$\eta_K^c(\boldsymbol{\tau}) = \|\boldsymbol{\tau} - \nabla u_h\|_K + \min\left\{\frac{h_K}{\pi}, \frac{1}{\kappa_K}\right\} \|\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}\|_K. \quad (13)$$

Moreover,

$$\eta_K^c(\boldsymbol{\tau}) \leq \sqrt{2} \min\{\eta_K^a(\boldsymbol{\tau}), \eta_K^b(\boldsymbol{\tau})\}. \quad (14)$$

Proof. Inequality (12) follows easily from the simple estimate

$$\eta_K^b(\boldsymbol{\tau}) \leq \|\boldsymbol{\tau} - \nabla u_h\|_K + \kappa_K^{-1} \|\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}\|_K.$$

Similarly, inequality (14) follows from the estimate

$$\|\boldsymbol{\tau} - \nabla u_h\|_K + \kappa_K^{-1} \|\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}\|_K \leq \sqrt{2} \eta_K^b(\boldsymbol{\tau}).$$

\square

Lemma 2 implies that we can replace $\min\{\eta_K^a(\boldsymbol{\tau}), \eta_K^b(\boldsymbol{\tau})\}$ by a simpler indicator $\eta_K^c(\boldsymbol{\tau})$ in (8) and the upper bound property still holds. On the other hand, indicator $\eta_K^c(\boldsymbol{\tau})$ is not as tight upper bound as $\min\{\eta_K^a(\boldsymbol{\tau}), \eta_K^b(\boldsymbol{\tau})\}$. It can overestimate it, but at most by a factor of $\sqrt{2}$.

3. Local flux reconstructions

All error indicators η^a , η^b , and η^c provide an upper bound on the energy norm of the error for a wide class of fluxes $\boldsymbol{\tau} \in \mathbf{W}$. However, an arbitrary choice of $\boldsymbol{\tau} \in \mathbf{W}$ would yield a large overestimation of the error. Therefore, the goal is to construct flux $\boldsymbol{\tau} \in \mathbf{W}$ that yields a tight bound. Tight bounds are provided by reconstructions of Ainsworth and Vejchodský [2] and Braess and Schöberl [5] and in this paper we compare their accuracy.

The reconstruction of Ainsworth and Vejchodský [2] is based on a fast algorithm to compute boundary fluxes on element edges. These boundary fluxes are computed by solving small so-called ‘topology’ systems of linear algebraic equations on patches of elements sharing a common vertex. Subsequently, the flux $\boldsymbol{\tau} \in \mathbf{W}$ is reconstructed element-by-element using explicit formulae that differ for small and large values of κh . The resulting error bound is locally efficient and robust with respect to both the mesh size h and the reaction coefficient κ [2] over the entire range of values of κh . For future reference, we denote this flux by $\boldsymbol{\tau}_h^{\text{AV}}$.

The reconstruction of Braess and Schöberl [5] is based on a solution of local problems on patches of elements around vertices of the triangulation. These local problems are formulated as mixed finite element problems and correspond to the minimization of the error bound localized to the patch with an equilibration constraint. Although this flux reconstruction was originally designed for pure diffusion problems, its generalization to the reaction-diffusion case is straightforward. However, this straightforward generalization does not yield good results for large values of κh as we will see below. For future reference, we denote this flux by $\boldsymbol{\tau}_h^{\text{BS}}$.

Both of these flux reconstructions have similar features. For example, in both cases the flux reconstruction is local and based on patches of elements sharing a common vertex. If $\kappa_K h_K$ is small, namely at most of order 1, then the flux $\boldsymbol{\tau}_h^{\text{AV}}$ lies in the Brezzi-Douglas-Marini space $\mathbf{BDM}^2(\mathcal{T}_h) = \{\boldsymbol{w}_h \in \mathbf{H}(\text{div}, \Omega_h) : \boldsymbol{w}_h|_K \in [P^2(K)]^2 \forall K \in \mathcal{T}_h\}$, while the flux $\boldsymbol{\tau}_h^{\text{BS}}$ lies in the Raviart-Thomas-Nédélec space $\mathbf{RTN}^1(\mathcal{T}_h) = \{\boldsymbol{w}_h \in \mathbf{H}(\text{div}, \Omega_h) : \boldsymbol{w}_h|_K \in [P^1(K)]^2 \oplus \boldsymbol{x}P^1(K) \forall K \in \mathcal{T}_h\}$. Spaces $\mathbf{BDM}^2(\mathcal{T}_h)$ and $\mathbf{RTN}^1(\mathcal{T}_h)$ are quite similar. They both contain piecewise quadratic vector fields and $\mathbf{RTN}^1(\mathcal{T}_h) \subset \mathbf{BDM}^2(\mathcal{T}_h)$. In addition, these flux reconstructions are exactly equilibrated, i.e. $\Pi_K f - \kappa_K^2 u_h + \text{div } \boldsymbol{\tau} = 0$ in all elements $K \in \mathcal{T}_h$ for both $\boldsymbol{\tau} = \boldsymbol{\tau}_h^{\text{AV}}$ and $\boldsymbol{\tau} = \boldsymbol{\tau}_h^{\text{BS}}$, provided $\kappa_K h_K$ is small. This means that in this case all three error indicators η_K^a , η_K^b , and η_K^c are actually equal for both $\boldsymbol{\tau}_h^{\text{AV}}$ and $\boldsymbol{\tau}_h^{\text{BS}}$. The situation is slightly different if $\kappa_K h_K$ is large, because then the reconstruction $\boldsymbol{\tau}_h^{\text{AV}}$ no longer satisfies the exact equilibration condition and it does not lie in $\mathbf{BDM}^2(\mathcal{T}_h)$ any more. Instead, it lies in $\mathbf{BDM}^2(\mathcal{T}_h^*)$, where \mathcal{T}_h^* is a certain special refinement of \mathcal{T}_h , and the employed error bound is η_K^b .

These similarities motivate our interest in the comparison of these two approaches. We compare them numerically on a couple of examples and find what reconstruction provides more accurate results. The second question is, what is the absolute accuracy of these local reconstructions and what is their potential for improvement. To answer this, we find the optimal flux reconstruction in the space $\mathbf{RTN}^1(\mathcal{T}_h)$. The optimal flux is obtained by a global minimization of the error bound under the weakest equilibration constraints.

4. Global flux reconstructions

In this section we present a procedure yielding the optimal flux reconstruction in a certain finite dimensional affine subspace $\mathbf{W}_h \subset \mathbf{W}$. The idea is to minimize the error bound (8) over \mathbf{W}_h . Since this error bound consists of a sum of error indicators and oscillation terms which are independent of $\boldsymbol{\tau}$ we minimize the sum of indicators only. The three error indicators we defined above correspond to the following three minimization problems:

$$\boldsymbol{\tau}_h^a = \arg \min_{\boldsymbol{\tau}_h \in \widetilde{\mathbf{W}}_h} \sum_{K \in \mathcal{T}_h} [\eta_K^a(\boldsymbol{\tau}_h)]^2, \quad (15)$$

$$\boldsymbol{\tau}_h^b = \arg \min_{\boldsymbol{\tau}_h \in \mathbf{W}_h} \sum_{K \in \mathcal{T}_h} [\eta_K^b(\boldsymbol{\tau}_h)]^2, \quad (16)$$

$$\boldsymbol{\tau}_h^c = \arg \min_{\boldsymbol{\tau}_h \in \widetilde{\mathbf{W}}_h} \sum_{K \in \mathcal{T}_h} [\eta_K^c(\boldsymbol{\tau}_h)]^2, \quad (17)$$

where $\widetilde{\mathbf{W}}_h = \{\boldsymbol{\tau}_h \in \mathbf{W}_h : \text{condition (5) holds for all } K \in \mathcal{T}_h\}$ is a subset of \mathbf{W}_h . Recalling the definitions (6) and (13) of η_K^a and η_K^c , we notice that the structure of problems (15) and (17) is the same. They are both constrained minimization problems and the only difference of indicators $\eta_K^a \in \widetilde{\mathbf{W}}_h$ and $\eta_K^c \in \widetilde{\mathbf{W}}_h$ is the constant multiple of the second term. On the other hand, minimization problem (16) is unconstrained and the indicator $\eta_K^b \in \mathbf{W}_h$ has a different structure.

Clearly, problem (16) is a quadratic minimization problem, but problems (15) and (17) are not quadratic. Since minimization of quadratic functionals is straightforward, we transform problems (15) and (17) such that they correspond to the minimization of a functional quadratic in $\boldsymbol{\tau}_h$. For example, in case (15), we use inequality

$$[A_K(\boldsymbol{\tau}_h) + B_K(\boldsymbol{\tau}_h)]^2 \leq \left(1 + \frac{1}{\xi_K}\right) A_K^2(\boldsymbol{\tau}_h) + (1 + \xi_K) B_K^2(\boldsymbol{\tau}_h),$$

where $A_K(\boldsymbol{\tau}_h) = \|\boldsymbol{\tau}_h - \nabla u_h\|_K$, $B_K(\boldsymbol{\tau}_h) = (h_K/\pi) \|\Pi_K f - \kappa_K^2 u_h + \text{div } \boldsymbol{\tau}_h\|_K$, and $\xi_K > 0$ is arbitrary. This inequality holds as equality if $\xi_K = A_K(\boldsymbol{\tau}_h)/B_K(\boldsymbol{\tau}_h)$. Thus, instead of minimizing the left-hand side of this inequality over $\boldsymbol{\tau}_h$, we equivalently minimize the right hand side over both $\xi_K > 0$ and $\boldsymbol{\tau}_h$. Note that the right-hand side is already quadratic in $\boldsymbol{\tau}_h$, but the nonlinear nature of the minimization problem

cannot be avoided and manifests itself in the nonlinear minimization with respect to ξ_K .

Using this approach, we reformulate all problems (15)–(17) to the minimization of the functional

$$J(\alpha_K, \beta_K, \boldsymbol{\tau}_h) = \sum_{K \in \mathcal{T}_h} \alpha_K \|\boldsymbol{\tau}_h - \nabla u_h\|_K^2 + \beta_K \|\Pi_K f - \kappa_K^2 u_h + \operatorname{div} \boldsymbol{\tau}_h\|_K^2, \quad (18)$$

where α_K and β_K are suitable constants defined for all elements $K \in \mathcal{T}_h$. For convenience, we use the notation $P^0(\mathcal{T}_h) = \{\xi \in L^1(\Omega_h) : \xi|_K = \xi_K \text{ is a constant } \forall K \in \mathcal{T}_h\}$. Problems (15)–(17), respectively, are then equivalent to:

$$(\boldsymbol{\tau}_h^a, \xi^a) = \arg \min_{\boldsymbol{\tau}_h \in \widetilde{\mathbf{W}}_h, \xi \in P^0(\mathcal{T}_h)} J(1 + \xi_K^{-1}, (1 + \xi_K)h_K^2/\pi^2, \boldsymbol{\tau}_h), \quad (19)$$

$$\boldsymbol{\tau}_h^b = \arg \min_{\boldsymbol{\tau}_h \in \mathbf{W}_h} J(1, \kappa_K^{-2}, \boldsymbol{\tau}_h), \quad (20)$$

$$(\boldsymbol{\tau}_h^c, \xi^c) = \arg \min_{\boldsymbol{\tau}_h \in \widetilde{\mathbf{W}}_h, \xi \in P^0(\mathcal{T}_h)} J(1 + \xi_K^{-1}, (1 + \xi_K) \min\{h_K^2/\pi^2, \kappa_K^{-2}\}, \boldsymbol{\tau}_h). \quad (21)$$

Note that in practice we solve problem (19) iteratively. We start with the natural choice $\xi_K \equiv 1$ for all $K \in \mathcal{T}_h$, fix it, and solve the quadratic minimization problem for $\boldsymbol{\tau}_h \in \widetilde{\mathbf{W}}_h$. Then we update ξ_K to $\xi_K = A_K(\boldsymbol{\tau}_h)/B_K(\boldsymbol{\tau}_h)$ for all $K \in \mathcal{T}_h$ and repeat the procedure until we find an (approximate) fixed point for ξ_K . The case of problem (21) is completely analogous.

Thus, for fixed ξ_K , both problems (19) and (21) are quadratic minimization problems for the functional (18) with suitable and fixed choices of constants α_K and β_K . Namely, $\alpha_K = 1 + \xi_K^{-1}$ and $\beta_K = (1 + \xi_K)h_K^2/\pi^2$ for problem (19) and $\alpha_K = 1 + \xi_K^{-1}$ and $\beta_K = (1 + \xi_K) \min\{h_K^2/\pi^2, \kappa_K^{-2}\}$ for problem (21). The constraint for these minimizations is the equilibration (5), see the definition of $\widetilde{\mathbf{W}}_h$. The solution of this constrained minimization problem can be obtained by solving the corresponding Euler-Lagrange equations: find $\boldsymbol{\tau}_h \in \mathbf{W}_h$ and $d_h \in P^0(\mathcal{T}_h)$ such that

$$\mathcal{B}^*(\boldsymbol{\tau}_h, \mathbf{w}_h) + \mathcal{Q}^*(d_h, \mathbf{w}_h) = \mathcal{F}^*(\mathbf{w}_h) \quad \forall \mathbf{w}_h \in \mathbf{W}_h, \quad (22)$$

$$-\mathcal{Q}^*(q_h, \boldsymbol{\tau}_h) = \mathcal{G}^*(q_h) \quad \forall q_h \in P^0(\mathcal{T}_h), \quad (23)$$

where

$$\mathcal{B}^*(\boldsymbol{\tau}_h, \mathbf{w}_h) = \sum_{K \in \mathcal{T}_h} \int_K (\alpha_K \boldsymbol{\tau}_h \cdot \mathbf{w}_h + \beta_K \operatorname{div} \boldsymbol{\tau}_h \operatorname{div} \mathbf{w}_h) \, d\mathbf{x},$$

$$\mathcal{Q}^*(d_h, \mathbf{w}_h) = \sum_{K \in \mathcal{T}_h} \int_K d_h \operatorname{div} \mathbf{w}_h \, d\mathbf{x},$$

$$\mathcal{F}^*(\mathbf{w}_h) = \sum_{K \in \mathcal{T}_h} \int_K (\alpha_K \nabla u_h \cdot \mathbf{w}_h - \beta_K (\Pi_K f - \kappa_K^2 u_h) \operatorname{div} \mathbf{w}_h) \, d\mathbf{x},$$

$$\mathcal{G}^*(q_h) = \sum_{K \in \mathcal{T}_h} \int_K (\Pi_K f - \kappa_K^2 u_h) q_h \, d\mathbf{x}.$$

Note that equality (23) corresponds to the equilibration constraint (5) and that d_h is the Lagrange multiplier. Consequently, if $\boldsymbol{\tau}_h \in \mathbf{W}_h$ solves (22)–(23) then it lies actually in $\widetilde{\mathbf{W}}_h$.

The case of the minimization problem (20) is even simpler. It is a quadratic minimization with no constraints. Therefore, its solution $\boldsymbol{\tau}_h \in \mathbf{W}_h$ is given by the corresponding Euler-Lagrange equations

$$\mathcal{B}^*(\boldsymbol{\tau}_h, \mathbf{w}_h) = \mathcal{F}^*(\mathbf{w}_h) \quad \forall \mathbf{w}_h \in \mathbf{W}_h, \quad (24)$$

where the constants α_K and β_K are 1 and κ_K^{-2} , respectively.

5. Numerical results

In this section, we consider two examples of reaction-diffusion problems. We solve them on a series of uniformly refined meshes and compute several error bounds of the form (8). In particular, we compute three error bounds η^a , η^b , and η^c , which are obtained from (4) by using indicators η_K^a , η_K^b , and η_K^c in place of η_K . In addition, we compute five different flux reconstructions. Namely, the local reconstructions $\boldsymbol{\tau}_h^{\text{AV}}$ and $\boldsymbol{\tau}_h^{\text{BS}}$ described in Section 3, and three global reconstructions $\boldsymbol{\tau}_h^a$, $\boldsymbol{\tau}_h^b$, and $\boldsymbol{\tau}_h^c$ in $\mathbf{W}_h = \mathbf{RTN}^1(\mathcal{T}_h) \cap \mathbf{W}$, see Section 4. Recall that reconstructions $\boldsymbol{\tau}_h^{\text{AV}}$ and $\boldsymbol{\tau}_h^{\text{BS}}$ are fully equilibrated and thus $\eta^a(\boldsymbol{\tau}_h^{\text{AV}}) = \eta^b(\boldsymbol{\tau}_h^{\text{AV}}) = \eta^c(\boldsymbol{\tau}_h^{\text{AV}})$ and $\eta^a(\boldsymbol{\tau}_h^{\text{BS}}) = \eta^b(\boldsymbol{\tau}_h^{\text{BS}}) = \eta^c(\boldsymbol{\tau}_h^{\text{BS}})$. For simplicity, we denote these two numbers by $\eta(\boldsymbol{\tau}_h^{\text{AV}})$ and $\eta(\boldsymbol{\tau}_h^{\text{BS}})$, respectively. Further, we use Lemma 2 to improve the error bound obtained by $\eta_K^c(\boldsymbol{\tau}_h^c)$. Once, we have computed $\boldsymbol{\tau}_h^c$, which is an expensive calculation, we can virtually for free evaluate the error bound

$$\eta^{\min}(\boldsymbol{\tau}_h^c) = \min\{\eta^a(\boldsymbol{\tau}_h^c), \eta^b(\boldsymbol{\tau}_h^c)\},$$

which is guaranteed to be less than or equal to $\eta_K^c(\boldsymbol{\tau}_h^c)$ and Theorem 1 implies that it is still an upper bound on the error. In order to compare the accuracy of these error bounds we use the index of effectivity $I_{\text{eff}} = \eta(\boldsymbol{\tau}_h)/\|u - u_h\|$, where u is the exact solution of problem (1) defined in Ω .

Example 1. Let us consider problem (1) in the domain of the shape of three quarters of a circular disk. Namely $\Omega = \{(r, \theta) : 0 \leq r < R \text{ and } \pi/2 < \theta < 2\pi\}$, where (r, θ) are the usual polar coordinates. We set $\Gamma_D = \partial\Omega$, $\Gamma_N = \emptyset$, and $f(r, \theta) = (32R^{-4/3}/9 + \kappa^2 r^{2/3} - \kappa^2 R^{-4/3} r^2) \sin(2\theta/3 - \pi/3)$. The exact solution to this problem $u(r, \theta) = (r^{2/3} - R^{-4/3} r^2) \sin(2\theta/3 - \pi/3)$ has a singularity at the re-entrant corner and we will use it to compute the energy norm $\|u - u_h\|$ of the error. For simplicity, we consider $R = 1$ and solve the problem for various constant values of κ .

The coarsest mesh we use is shown in Figure 1 (left). We then uniformly refine this mesh several times and compute the indices of effectivity for the above described error bounds on this sequence of meshes. Figure 2 (left) presents these results for

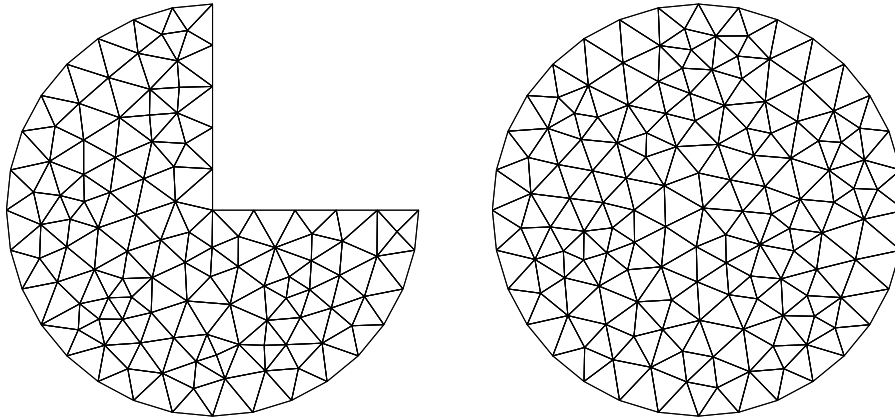


Figure 1: Domains and the coarsest meshes for Examples 1 (left) and 2 (right).

$\kappa = 100$. These results confirm that all these error bounds behave robustly with respect to the mesh size, although the case $\kappa = 100$ seems to be difficult for error bounds of this type and several of them yield indices of effectivity up to 5.

Figure 2 (right) shows how these indices of effectivity vary with κ , provided the mesh is fixed. We have chosen two times refined initial mesh. We observe that not all the error bounds provide robust bounds over this range of κ . Estimators $\eta(\boldsymbol{\tau}_h^{\text{BS}})$ and $\eta^a(\boldsymbol{\tau}_h^a)$ overestimate the error hugely if κ is large ($\kappa \geq 100$ in this case). This is not too surprising, because these two error bounds are not designed to be robust in the singularly perturbed case. On the other hand, for small values of κ (below 100) all error bounds provide very accurate results with indices of effectivity below 1.2. Only the bound $\eta(\boldsymbol{\tau}_h^{\text{AV}})$ yields indices of effectivity around 1.7.

Example 2. Let Ω be a unit disk, $f = 1$, and homogeneous Dirichlet boundary conditions be prescribed on the boundary of Ω . Then, the exact solution of problem (1) is $u = (1 - r^2)/4$ for $\kappa = 0$ and $u = \kappa^{-2}(1 - I_0(\kappa r)/I_0(\kappa))$ for $\kappa > 0$. Here, $r^2 = x^2 + y^2$ and I_0 stands for the modified Bessel function of the first kind.

As in Example 1, we solve this problem on a series of uniformly refined meshes, where the coarsest mesh is presented in Figure 1 (right). Figure 3 presents the results in the same manner as Figure 2. Conclusions are basically the same as for Example 1. A difference is that in this example all error bounds provide consistently better results than in Example 1. The reason probably is that the exact solution in this example has no singularity and that the right-hand side f is constant and thus, there are no quadrature errors and the oscillation term vanishes.

If κ is small (below 100) then all error bounds yield almost exact results. An exception is $\eta(\boldsymbol{\tau}_h^{\text{AV}})$ which overestimates the error by about 7% with a worse accuracy already for $\kappa = 10$. On the other hand if κ is large (above 100) then the local bound $\eta(\boldsymbol{\tau}_h^{\text{BS}})$ and the global bound $\eta^a(\boldsymbol{\tau}_h^a)$ overestimate the error hugely. However, all the other error bounds provide almost exact results. The intermediate range of values of κ around 100 seems to be problematic for all considered error bounds, because they all exhibit the least accurate values there.

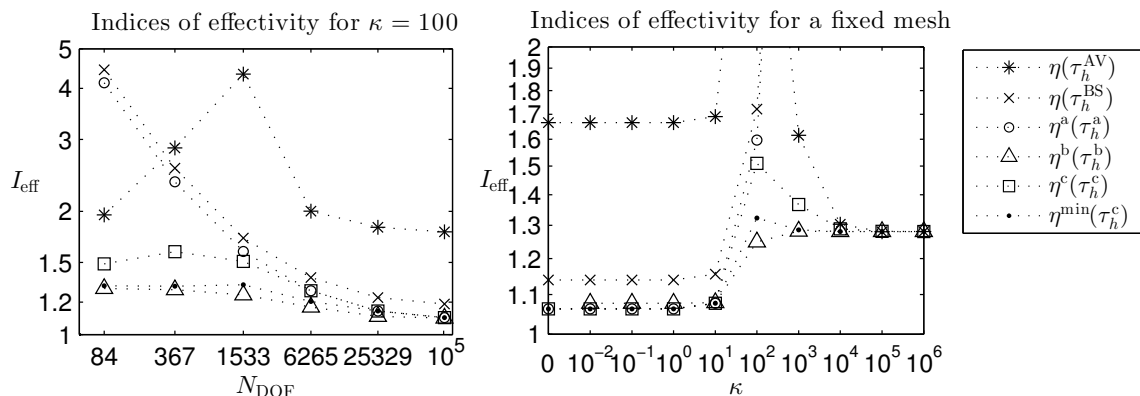


Figure 2: Results of Examples 1. The left panel shows the variations of the index of effectivity for various error bounds on a sequence of uniformly refined meshes for $\kappa = 100$. The mesh sizes h for these meshes are approximately 0.24, 0.12, 0.060, 0.030, 0.015, 0.0075, respectively. The right panel presents their variation with respect to κ on the mesh with $N_{\text{DOF}} = 1533$ ($h \approx 0.060$), i.e. the two times refined the initial mesh.

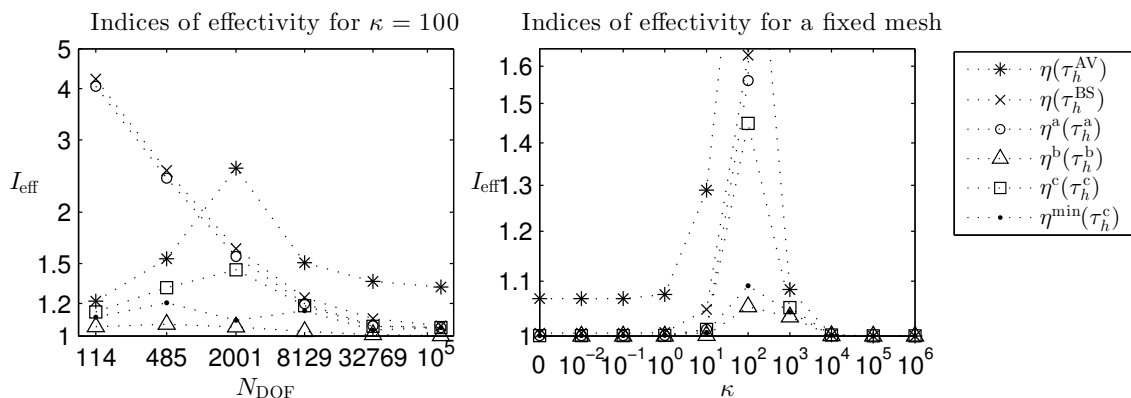


Figure 3: Results of Examples 2. The left panel shows the variations of the index of effectivity for various error bounds on a sequence of uniformly refined meshes for $\kappa = 100$. The mesh sizes h for these meshes are approximately 0.24, 0.12, 0.062, 0.031, 0.016, 0.0078, respectively. The right panel presents their variation with respect to κ on the mesh with $N_{\text{DOF}} = 2001$ ($h \approx 0.062$), i.e. the two times refined the initial mesh.

6. Conclusions

We have compared the accuracy of two local flux reconstructions and assessed their accuracy with respect to optimal reconstructions computed as global minimization problems. We observe that both the locally computed error bounds provide good accuracy if κh is smaller than approximately 1/2. However, the bound $\eta(\tau_h^{\text{BS}})$ provides results close to the optimal values computed by the global minimization and

performs considerably better than $\eta(\boldsymbol{\tau}_h^{\text{AV}})$. On the other hand, if κh is larger than approximately 50 then the bound $\eta(\boldsymbol{\tau}_h^{\text{BS}})$ overestimates the true error unacceptably. The reason is the unnatural form of the error bound and too restrictive equilibration of $\boldsymbol{\tau}_h^{\text{BS}}$. Similarly, even the globally computed bound $\eta^{\text{a}}(\boldsymbol{\tau}_h^{\text{a}})$ overestimates the error unacceptably. Nevertheless, all the other error bounds provide accurate results. Namely, the local flux reconstruction $\boldsymbol{\tau}_h^{\text{AV}}$ yields practically as accurate results as the globally computed optimal reconstructions. The intermediate values of κh seem to be problematic for the accuracy of error bounds of the considered type. Although all error bounds provide acceptable results for these values of κh , their accuracy is worse and sometimes considerably worse than their accuracy for other values of κh . (Bound $\eta^{\text{b}}(\boldsymbol{\tau}_h^{\text{b}})$ in Example 1 being the only exception in the provided examples.)

In general, comparing the two locally computed error bounds, we may conclude that $\eta(\boldsymbol{\tau}_h^{\text{BS}})$ is very accurate and yields close to optimal results for κh small. However, if κh is large then $\eta(\boldsymbol{\tau}_h^{\text{BS}})$ fails. The second locally constructed error bound $\eta(\boldsymbol{\tau}_h^{\text{AV}})$ is less accurate for small values of κh , but still provides acceptable values. For large values of κh it gives nearly optimal results.

The secondary conclusion we can draw from the performed experiments, concerns the globally computed reconstructions and the three possible forms of the error bounds. The form η^{a} cannot be recommended in general, because it provides accurate results for small values of κh only. The form η^{b} provides accurate results over the whole range of values of $\kappa h > 0$. Even more, it provides the best results except for cases with small κh , where it is only slightly worse than η^{a} . The disadvantage of η^{b} is the fact that it is undefined in the important case $\kappa = 0$. Therefore, we can recommend to use η^{c} as a robust solution. The bound η^{c} and especially its improved variant η^{min} provides results that are close to the best in all cases.

The obtained results suggest several directions for future investigations. First, there is a potential for further improvements of the bound $\eta(\boldsymbol{\tau}_h^{\text{AV}})$, which is not as accurate as it could be for small and intermediate values of κh . Second, the bound $\eta(\boldsymbol{\tau}_h^{\text{AV}})$ is almost optimal for large values of κh , but this flux reconstruction is constructed on the refined mesh \mathcal{T}_h^* . However, the performance of the global flux reconstructions $\boldsymbol{\tau}_h^{\text{b}}$ and $\boldsymbol{\tau}_h^{\text{c}}$ clearly shows that a robust reconstruction is possible even on the original mesh \mathcal{T}_h . Therefore, we may try to simplify the construction of $\boldsymbol{\tau}_h^{\text{AV}}$ for large values of κh and define it on \mathcal{T}_h only while keeping its robust and accurate performance. Third, the bound $\eta(\boldsymbol{\tau}_h^{\text{BS}})$ can be improved and redefined in such a way that it is robust and accurate even in the singularly perturbed case.

Finally, let us point out that the presented error bounds estimate the error $\tilde{u} - u_h$, which includes the discretization error, quadrature errors, round-off errors, and the error of the solver of linear algebraic equations. However, these error bounds ignore the domain approximation error $u - \tilde{u}$. Therefore, they could theoretically underestimate the total error $u - u_h$ in the case of large domain approximation error. Both the presented examples exhibit nonzero domain approximation error, but the used meshes seem to approximate the exact domain Ω well, because we do not observe any indices of effectivity below zero.

Acknowledgements

This work has been supported by grant No. P101/14-02067S of the Czech Science Foundation and by RVO 67985840.

References

- [1] Ainsworth, M. and Vejchodský, T.: Fully computable robust a posteriori error bounds for singularly perturbed reaction–diffusion problems. *Numer. Math.* **119** (2011), 219–243.
- [2] Ainsworth, M. and Vejchodský, T.: Robust error bounds for finite element approximation of reaction-diffusion problems with non-constant reaction coefficient in arbitrary space dimension. *Comput. Methods Appl. Mech. Engrg.* **281** (2014), 184–199.
- [3] Aubin, J. P. and Burchard, H. G.: Some aspects of the method of the hypercircle applied to elliptic variational problems. In: *Numerical Solution of Partial Differential Equations, II (SYNSPADE 1970) (Proc. Sympos., Univ. of Maryland, College Park, Md., 1970)*, pp. 1–67. Academic Press, New York, 1971.
- [4] Braess, D.: *Finite elements: Theory, fast solvers, and applications in elasticity theory*. Cambridge University Press, Cambridge, 2007, 3rd edn.
- [5] Braess, D. and Schöberl, J.: Equilibrated residual error estimator for edge elements. *Math. Comp.* **77** (2008), 651–672.
- [6] Cheddadi, I., Fučík, R., Prieto, M.I., and Vohralík, M.: Guaranteed and robust a posteriori error estimates for singularly perturbed reaction–diffusion problems. *M2AN Math. Model. Numer. Anal.* **43** (2009), 867–888.
- [7] Destuynder, P. and Métivet, B.: Explicit error bounds in a conforming finite element method. *Math. Comp.* **68** (1999), 1379–1396.
- [8] Haslinger, J. and Hlaváček, I.: Convergence of a finite element method based on the dual variational formulation. *Apl. Mat.* **21** (1976), 43–65.
- [9] Kelly, D. W.: 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.
- [10] Křížek, M.: Conforming equilibrium finite element methods for some elliptic plane problems. *RAIRO Anal. Numér.* **17** (1983), 35–65.
- [11] Kunert, G.: A posterior H^1 error estimation for a singularly perturbed reaction diffusion problem on anisotropic meshes. *IMA J. Numer. Anal.* **25** (2005), 408–428.

- [12] Ladevèze, P. and Leguillon, D.: Error estimate procedure in the finite element method and applications. *SIAM J. Numer. Anal.* **20** (1983), 485–509.
- [13] Parés, N., Santos, H., and Díez, P.: Guaranteed energy error bounds for the Poisson equation using a flux-free approach: solving the local problems in subdomains. *Internat. J. Numer. Methods Engrg.* **79** (2009), 1203–1244.
- [14] Payne, L. E. and Weinberger, H. F.: An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.* **5** (1960), 286–292 (1960).
- [15] Prager, W. and Synge, J. L.: Approximations in elasticity based on the concept of function space. *Quart. Appl. Math.* **5** (1947), 241–269.
- [16] Repin, S.: *A posteriori estimates for partial differential equations, Radon Series on Computational and Applied Mathematics*, vol. 4. de Gruyter, Berlin, 2008.
- [17] Repin, S. and Sauter, S.: Functional a posteriori estimates for the reaction-diffusion problem. *C. R. Math. Acad. Sci. Paris* **343** (2006), 349–354.
- [18] Šebestová, I. and Vejchodský, T.: Two-sided bounds for eigenvalues of differential operators with applications to Friedrichs, Poincaré, trace, and similar constants. *SIAM J. Numer. Anal.* **52** (2014), 308–329.
- [19] Synge, J. L.: *The hypercircle in mathematical physics: a method for the approximate solution of boundary value problems*. Cambridge University Press, New York, 1957.
- [20] Vejchodský, T.: Guaranteed and locally computable a posteriori error estimate. *IMA J. Numer. Anal.* **26** (2006), 525–540.
- [21] Vejchodský, T.: Complementarity based a posteriori error estimates and their properties. *Math. Comput. Simulation* **82** (2012), 2033–2046.
- [22] Verfürth, R.: A note on constant-free a posteriori error estimates. *SIAM J. Numer. Anal.* **47** (2009), 3180–3194.
- [23] de Veubeke, B. F.: Displacement and equilibrium models in the finite element method. In: O. Zienkiewicz and G. Hollister (Eds.), *Stress Analysis*, pp. 145–197. Wiley, London, 1965.
- [24] Zhang, B., Chen, S., and Zhao, J.: Guaranteed a posteriori error estimates for nonconforming finite element approximations to a singularly perturbed reaction–diffusion problem. *Appl. Numer. Math.* **94** (2015), 1–15.