JYVÄSKYLÄ STUDIES IN COMPUTING

190

Immanuel Anjam

A Posteriori Error Control for Maxwell and Elliptic Type Problems

JYVÄSKYLÄN YLIOPISTO

JYVÄSKYLÄ STUDIES IN COMPUTING 190

Immanuel Anjam

A Posteriori Error Control for Maxwell and Elliptic Type Problems

Esitetään Jyväskylän yliopiston informaatioteknologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Agora-rakennuksen auditoriossa 2
kesäkuun 27. päivänä 2014 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Information Technology of the University of Jyväskylä,
in building Agora, Auditorium 2, on June 27, 2014 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2014

A Posteriori Error Control for Maxwell and Elliptic Type Problems

JYVÄSKYLÄ STUDIES IN COMPUTING 190

Immanuel Anjam

A Posteriori Error Control for Maxwell and Elliptic Type Problems

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2014

Editors
Timo Männikkö
Department of Mathematical Information Technology, University of Jyväskylä
Pekka Olsbo, Ville Korkiakangas
Publishing Unit, University Library of Jyväskylä

URN:ISBN:978-951-39-5731-5
ISBN 978-951-39-5731-5 (PDF)

ISBN 978-951-39-5730-8 (nid.)
ISSN 1456-5390

Jyväskylä University Printing House, Jyväskylä 2014

ABSTRACT

Anjam, Immanuel
A posteriori error control for Maxwell and elliptic type problems
Jyväskylä: University of Jyväskylä, 2014, 74 p. (+included articles)
(Jyväskylä Studies in Computing
ISSN 1456-5390; **190**)
ISBN 978-951-39-5730-8 (nid.)
ISBN 978-951-39-5731-5 (PDF)
Finnish summary
Diss.

In the included papers of this thesis we research functional type a posteriori error estimates and indicators for the eddy-current problem, Stokes problem, and the diffusion problem. The summary of this thesis concerns only the eddy-current problem. This is a second order partial differential equation derived from the Maxwell equations. We derive and numerically test the classical functional a posteriori error estimates, and a new error equality for mixed approximations. All presented error estimates in this thesis hold for all conforming approximations, and only contain global constants. The error equality does not even contain any constants. We also study the radiuses of the solution sets generated by indeterminate right hand side, and indeterminate material parameters. Computable quantities for practical simulations are presented and numerically tested. The effect of indeterminate material parameters on error indication is also investigated by computational means.

Keywords: functional a posteriori error control, error estimate, error indicator, error equality, elliptic problem, Maxwell type problem, eddy-current problem

Author

Immanuel Anjam

Department of Mathematical Information Technology
University of Jyväskylä
Finland

Supervisors

Prof. Pekka Neittaanmäki

Department of Mathematical Information Technology
University of Jyväskylä
Finland

Prof. Sergey Repin

Steklov Institute, St. Petersburg department
Russian Academy of Sciences
Russia

Department of Mathematical Information Technology
University of Jyväskylä
Finland

Reviewers

Prof. Karl-Josef Witsch

Department of Mathematics
University of Duisburg-Essen
Germany

Dr. Tatiana Samrowski

School of Engineering
Zürich University of Applied Sciences
Switzerland

Institute of Mathematics
Zürich University
Switzerland

Opponent

Prof. Johannes Kraus

Department of Mathematics
University of Duisburg-Essen
Germany

Johann Radon Institute for
Computational and Applied Mathematics
Austria

ACKNOWLEDGEMENTS

I am greatly indebted to my supervisors Pekka Neittaanmäki and Sergey Repin for introducing me to the world of science, and for excellent supervision, which began already in 2007 during my bachelor studies. I also thank my supervisors for time spent outside of scientific work.

I also want to acknowledge the COMAS Graduate School, KAUTE foundation, Emil Aaltonen foundation and Väisälä foundation (of the Finnish Academy of Science and Letters) for funding my PhD studies.

I would like to express my thanks for the reviewers Karl-Josef Witsch and Tatiana Samrowski for significantly improving the quality of the thesis during the review process.

I thank all the members of the research group "Reliable Methods for Computer Simulation" for interesting discussions we had both about scientific and non-scientific matters. I want to especially thank Olli Mali for his huge involvement in teaching me numerics during my bachelor and master studies, and for support during my PhD studies.

I also thank Dirk Pauly for support during my PhD studies, and especially for inviting me to the Fakultät für Mathematik of the Universität Duisburg-Essen for the academic year 2012-2013. I thank Karl-Josef Witsch for lecturing an excellent course about partial differential equations during my visit.

Lastly I thank my friends and family who certainly had nothing to do with my academic adventures, but who contributed to this journey by dragging me away from scientific work ever so often.

Immanuel Anjam
Jyväskylä, May 2014

LIST OF FIGURES

LIST OF TABLES

CONTENTS

LIST OF INCLUDED ARTICLES

PI I. Anjam, O. Mali, A. Muzalevsky, P. Neittaanmäki, and S. Repin. A posteriori error estimates for a Maxwell type problem. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **24**(5):395–408, 2009.

PII I. Anjam, M. Nokka, and S. Repin. On a posteriori error bounds for approximations of the generalized Stokes problem generated by the Uzawa algorithm. *Russian Journal of Numerical Analysis and Mathematical Modelling*, **27**(4):321–338, 2012.

PIII I. Anjam and D. Pauly. Functional a posteriori error equalities for conforming mixed approximations of elliptic problems. *Reports of the Department of Mathematical Information Technology, University of Jyväskylä, Series B. Scientific Computing, No. B. 2/2014, ISBN 978-951-39-5643-1, ISSN 1456-436X*, 2014.

PIV I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin. New indicators of approximation errors for problems in continuum mechanics. *J. C. F. Pereira, A. Sequeira and J. M. C. Pereira (Eds.), Proceedings of the V European Conference on Computational Fluid Dynamics ECCOMAS CFD 2010, Lisbon, Portugal,14-17 June 2010, ISBN: 978-989-96778-1-4*, 2010.

PV I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin. On the reliability of error indication methods for problems with uncertain data. *A. Cangiani, R.L. Davidchack, E. Georgoulis, A.N. Gorban, J. Levesley, and M.V. Tretyakov (Eds.), Numerical Mathematics and Advanced Applications 2011*, 811–819, 2013.

PVI I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin. A unified approach to measuring accuracy of error indicators. *W. Fitzgibbon, Yu. A. Kuznetsov, P. Neittaanmaki, and O. Pironneau (Eds.), Modeling, Simulation and Optimization for Science and Technology (to appear)*, 2014.

1 INTRODUCTION

The equations governing various phenomena in nature are partial differential equations (PDEs). Also various problems arising in industrial optimization include systems of PDEs. For example, the behavior of fluids, the distribution of heat in some material, and the behavior of electromagnetic fields are modeled using partial differential equations. Even calculating weather forecasts is a problem where solving a complicated system of PDEs plays an essential part.

Solutions of partial differential equations can be calculated exactly only in some special cases. These special cases are usually artificial and non-realistic. Typically, solutions of PDEs have to be calculated numerically with computers. Popular numerical methods include, i.e., the finite element method (FEM) [6, 8, 26, 29], the finite difference method (FDM) [27] and the finite volume method (FVM) [10]. Solving a PDE numerically results in an *approximate* solution. In some cases these approximate solutions might coincide with the corresponding exact solutions, but in general this is hardly the case. Usually approximate solutions always include some amount of error. In order for a numerical method to be verified as reliable, there must be a way to measure the error of the approximate solutions it produces.

There are two different ways to approach the problem of error measurement. One can measure the error *a priori*, before the approximate solutions are computed. A priori methods justify the applied numerical method, and often provide some information over the efficiency of the method. However, they do not say anything specific about the error of a particular approximate solution.

The second way of measuring the error is called *a posteriori*, where the error is measured after the approximate solutions are computed. The goal of a posteriori error control is to measure the error between the approximate solution and the exact solution. An a posteriori *error estimate* approximates the total error of the numerical solution, and an *error indicator* approximates the error distribution in the domain where the PDE is solved. Because the exact solution is unknown, the values given by some error measurement method are usually *approximations of the error*. Just like the approximation of a PDE can coincide with the exact solution, in some rare cases the approximation of the error might coincide with the exact error. However, this is not something that can not be expected with classical error measurement methods. That being said, a new result exposed in this thesis shows that for a certain class of PDEs the exact error is known even though the exact solution is not known.

There are several types of a posteriori error estimates and indicators: the residual method [3–5, 25, 26], equilibration based methods [7], and gradient averaging type methods [47, 48]. In this thesis, we research the *functional type* a posteriori error estimates and indicators invented by Prof. Sergey Repin [22, 31, 41]. These type estimates have several good properties. They are independent of the method which is used to produce approximate solutions. The only restriction is that the solution must belong to the correct function space, and satisfy the given boundary conditions exactly. In mathematical terms one would say that the approximation must be *conforming*. Also, these estimates contain only global constants. This means that the constants are not related to any function in the PDE, or the mesh used to discretize the domain. In comparison, for example the residual method is in many cases based in a very strong Galerkin orthogonality assumption, and also contains constants which are dependent on the mesh discretization.

The Maxwell equations is a first order system of four partial differential equations which govern the theory of electromagnetism. These equations have numerous practical applications [15, 26]. These include designing adapters, laser resonators, dynamos, radars, optics, etc. The Maxwell equations are rather relevant also in information and telecommunication technology: antennas are widely used electromagnetic applications.

In the summary of this thesis, we consider a second order PDE closely related to the Maxwell equations. In literature, this problem is often referred to as the *eddy-current* problem. In Chapter 2, we list notation and mathematical results used in the thesis. We also expose the model problem in this chapter, deriving it from the Maxwell equations. Chapter 3 is concentrated in functional a posteriori error estimates for the approximations of eddy-current equations. Also an error equality for combined error of mixed approximations is exposed in this chapter. In Chapter 4 we do an analysis of the effects of indeterminacy in the problem data.

Many of the topics of this thesis were suggested by the two supervisors Prof. Pekka Neittaanmäki and Prof. Sergey Repin, and are based on their previous research on the subject of a posteriori error control. The error equality of Section 3.4, and its use in analyzing indeterminate data in Sections 4.1 and 4.2 is original work of the author.

Of the included papers [PI] and [PIII] are directly related to this thesis, and the post-processing technique presented in the conference paper [PIV] is also derived and numerically tested in the context of the model problem of this thesis. The conference papers [PV] and [PVI] consider the reliability of error indication, and part of this analysis is repeated for the model problem of this thesis. The paper [PII] considers the generalized Stokes problem in its classical formulation, and is out of the main line of this thesis. However, the contents of this paper is

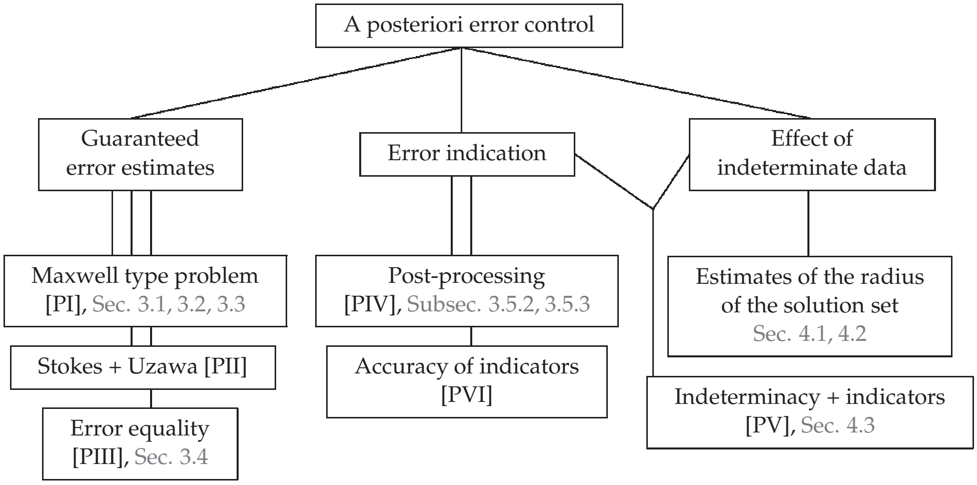

A posteriori error control
Guaranteed error estimates
Error indication
Effect of indeterminate data
Maxwell type problem [PI], Sec. 3.1, 3.2, 3.3
Post-processing [PIV], Subsec. 3.5.2, 3.5.3
Estimates of the radius of the solution set Sec. 4.1, 4.2
Stokes + Uzawa [PII]
Accuracy of indicators [PVI]
Indeterminacy + indicators [PV], Sec. 4.3
Error equality [PIII], Sec. 3.4

FIGURE 1 Structure of the thesis.

related to this thesis by the fact that the so-called vorticity-velocity-pressure formulation of the Stokes problem is very close to the model problem considered in this thesis.

Different parts of the summary of this thesis are related to the included papers according to the diagram in Figure 1. Results which are exposed in the included papers, or in other publications, will be highlighted accordingly.

Author's contribution to the included articles

[PI]: The estimates were already derived in earlier publications of the supervisors in [30, 43] and separately by A. Hannukainen in [12]. The author did the numerical results. The implementation of finite element solvers was done in close collaboration with co-author O. Mali. Co-author A. Muzalevsky also provided an implementation for comparison and verification of results.

[PII]: In this paper we exposed new type functional a posteriori error estimates specifically meant for approximations generated by the Uzawa algorithm. The mathematical results were done in close collaboration with co-authors M. Nokka and S. Repin. No numerical results were exposed in this article.

[PIII]: The equality presented in this paper is original research of the author. D. Pauly helped in writing the general section of the paper. The author did the numerical results.

[PIV]: The idea of edge-averaging and post-processing using the functional error majorants came from numerous discussions with the supervisors. The author carried out the necessary computations and did the numerical results.

[PV, PVI]: Comparing various error indication methods in the case of indeterminacy was done in close collaboration with the supervisors. For these articles the author implemented all the different error indication methods, and came up with an efficient way to test the effect of indeterminacy on error indication.

The author used solely `MATLAB` [23] for all numerical results, doing all the implementations of, i.e., finite elements by himself in order to have complete control over simulations. The meshes in 2d were handled by the functions provided by the toolbox `PDEtool`, but meshes in 3d were handled by code created by the author. In some cases some `BASH` code was used to help with running big simulations on computer clusters running on `Unix`-based operating systems.

The finite element implementations (for the primal and dual problems of the diffusion, reaction-diffusion, eddy-current, and Stokes problems) programmed by the author during the course of PhD studies include

- linear Courant element in 2d and 3d
- quadratic Courant element in 2d
- linear Raviart-Thomas element in 2d and 3d
- linear tensor-valued Raviart-Thomas element in 2d
- linear Nédélec element of the first family in 2d and 3d
- linear tensor-valued Nédélec element of the first family in 2d
- the scalar and vector valued MINI-element in 2d

In the majority of the finite-element solvers the `for`-loop over elements was replaced by vectorized operations creating surprisingly readable and fast solvers. The vectorization was done in the way presented in [39]. Also various averaging and post-processing techniques were implemented in both 2d and 3d for computationally cheap ways to obtain approximations of the dual variables. In addition to implementing functional type a posteriori error estimates, the author also implemented various other error indicators for the diffusion problem in order to compare their performance in the articles [PV, PVI].

2 MATHEMATICAL BACKGROUND AND THE MODEL PROBLEM

In this chapter we shortly describe the notation and several important results we will be using throughout this document. We also derive our model problem from the Maxwell equations.

We denote by $\mathbb{R}^d$ the space of d-dimensional real valued vectors, and by $\mathbb{M}^{d\times d}$ the space of real valued second order tensors ($d \geq 2$). The inner products of $a, b \in \mathbb{R}^d$ and $\mathbf{A}, \mathbf{B} \in \mathbb{M}^{d\times d}$ are defined as

$$a \cdot b := \sum_{i=1}^{d} a_i b_i \qquad \text{and} \qquad \mathbf{A} : \mathbf{B} := \sum_{i=1}^{d}\sum_{j=1}^{d} \mathbf{A}_{ij}\mathbf{B}_{ij}.$$

These inner products define the norms

$$\|a\|_{\mathbb{R}^d} := (a \cdot a)^{1/2} \qquad \text{and} \qquad \|\mathbf{A}\|_{\mathbb{M}} := (\mathbf{A} : \mathbf{A})^{1/2}.$$

Throughout this thesis, $\Omega \in \mathbb{R}^d$ denotes a bounded domain with Lipschitz continuous boundary $\partial\Omega$. Let Γ_D be a relative open subset of the boundary $\partial\Omega$, and $\Gamma_N := \partial\Omega \backslash \overline{\Gamma}_D$. A point in the domain Ω is denoted by $x = (x_1, x_2, \ldots, x_d)^T$, where T denotes the transpose.

In the scope of numerical results, we will also assume Ω to be a polyhedral domain in order for the mesh discretization to be exact at the boundary $\partial\Omega$. The reason for this is that the presented estimates require the approximation to be conforming, i.e., it must belong to the correct Sobolev space, and the boundary conditions must be satisfied exactly.

These requirements for Ω will not be repeated, and additional requirements will be separately emphasized if needed.

Spaces of square integrable functions

We define the following inner product and norm for the functions $f, g : \Omega \to \mathbb{R}$:

$$\begin{aligned}(f,g)_{L_2(\Omega)} &:= \int_\Omega fg \,\mathrm{dx},\\ \|f\|^2_{L_2(\Omega)} &:= (f,f)_{L_2(\Omega)} = \int_\Omega |f|^2 \,\mathrm{dx}.\end{aligned}$$

We refer to this inner product and norm as L^2-inner product and -norm, respectively. The space of real valued L^2-functions is then defined as

$$L^2(\Omega) := \{f : \Omega \to \mathbb{R} \mid \|f\|_{L^2(\Omega)} < \infty\},$$

which is a Hilbert space. A vector function $v : \Omega \to \mathbb{R}^d$ is said to belong to $L^2(\Omega, \mathbb{R}^d)$, if all its components $v_i \in L^2(\Omega)$ for all i. More formally, we define the following inner product and norm for vector functions v and w:

$$\begin{aligned}(v,w)_{L^2(\Omega,\mathbb{R}^d)} &:= \int_\Omega v \cdot w \,\mathrm{dx},\\ \|v\|^2_{L^2(\Omega,\mathbb{R}^d)} &:= (v,v)_{L^2(\Omega,\mathbb{R}^d)} = \int_\Omega |v|^2 \,\mathrm{dx}.\end{aligned}$$

With this inner product and corresponding norm, we define the L^2-space

$$L^2(\Omega, \mathbb{R}^d) := \{v : \Omega \to \mathbb{R}^d \mid \|v\|_{L^2(\Omega,\mathbb{R}^d)} < \infty\}.$$

For the rest of the thesis we will drop the subindices denoting spaces, and denote by $(\cdot,\cdot)_\omega$ and $\|\cdot\|_\omega$ the L^2-inner products and -norms of scalar and vector functions in $\omega \subset \Omega$. If $\omega = \Omega$, we will omit the domain from the subindex.

From now on, the subindex will be used for denoting *weighed* inner products and norms. For example, for vector valued functions v and w, we denote the δ-weighed L^2-inner product and -norm by

$$(v,w)_{\omega,\delta} := \int_\omega \delta v \cdot w \,\mathrm{dx}, \tag{2.1}$$

$$\|v\|^2_{\omega,\delta} := (v,v)_{\omega,\delta} = \int_\omega \delta v \cdot v \,\mathrm{dx}, \tag{2.2}$$

where $\delta \in L^\infty(\Omega, \mathbb{M}^{d\times d})$ is uniformly positive definite, real, and symmetric.

L^∞ denotes the space of *essentially bounded functions*. For a scalar function f, the L^∞-norm is defined as

$$\|f\|_{L^\infty(\Omega)} := \operatorname*{ess\,sup}_{x\in\Omega} |f(x)| \,.$$

The space of scalar L^∞-functions is then

$$L^\infty(\Omega) := \{f : \Omega \to \mathbb{R} \mid \|f\|_{L^\infty(\Omega)} < \infty\}.$$

We say that a vector function v belongs to $L^\infty(\Omega, \mathbb{R}^d)$ if all of its components belong to $L^\infty(\Omega)$.

Useful inequalities

We will use the following inequalities (see, e.g., [18]). The Cauchy-Schwarz inequality (a special case of the Hölder inequality) in its integral form reads

$$(f,g) \leq \|f\|\,\|g\| \tag{2.3}$$

for scalar functions $f,g \in L^2(\Omega)$. The discrete inequality reads

$$|a \cdot b| \leq \|a\|_{\mathbb{R}^d}\|b\|_{\mathbb{R}^d} \tag{2.4}$$

for $a,b \in \mathbb{R}^d$. The triangle inequality for scalar functions f and g is called the Minkowski inequality, and it reads

$$\|f+g\| \leq \|f\| + \|g\|. \tag{2.5}$$

The Young's inequality for scalars a and b reads

$$ab \leq \frac{1}{2\delta}a^2 + \frac{\delta}{2}b^2, \tag{2.6}$$

and it holds for all $\delta > 0$. Inequalities like (2.3) and (2.5) hold also for vector functions from $L^2(\Omega,\mathbb{R}^d)$.

Differential operators

Let f and v be smooth scalar and vector valued functions, respectively. For a scalar function f we define the gradient by

$$\nabla f := \begin{pmatrix} \partial_1 f \\ \partial_2 f \\ \vdots \\ \partial_d f \end{pmatrix} \in \mathbb{R}^d.$$

The divergence of v is defined as

$$\operatorname{div} v := \sum_{i=1}^{d} \partial_i v_i \in \mathbb{R}.$$

Note that the Laplace operator Δ can be represented as div∇.

The rotation operators differ depending on the dimension. If $d = 3$, the rotation is defined by

$$\operatorname{curl} v := \begin{pmatrix} \partial_2 v_3 - \partial_3 v_2, \\ \partial_3 v_1 - \partial_1 v_3, \\ \partial_1 v_2 - \partial_2 v_1 \end{pmatrix} \in \mathbb{R}^3.$$

If $d = 2$, we have two operators describing rotation, one for a scalar function f and one for a vector function v:

$$\underline{\operatorname{curl}}\, f := \begin{pmatrix} \partial_2 f \\ -\partial_1 f \end{pmatrix} \in \mathbb{R}^2,$$
$$\operatorname{curl} v := \partial_1 v_2 - \partial_2 v_1 \in \mathbb{R}.$$

The rotation operator $\underline{\mathrm{curl}}$ is often called "the co-gradient", and denoted by $\nabla^{\perp}$ in literature, since

$$\underline{\mathrm{curl}} f = \mathbf{R}\nabla f, \qquad \text{where} \qquad \mathbf{R} := \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Also, if $d = 2$, we have $\operatorname{curl} v = \operatorname{div} \mathbf{R} v$ and $\operatorname{curl} \mathbf{R} v = -\operatorname{div} v$. Moreover, $\operatorname{curl} \underline{\mathrm{curl}} = \operatorname{curl} \mathbf{R}\nabla = -\operatorname{div}\nabla = -\Delta$.

Sobolev spaces

From now on, differentials will be considered in the weak sense. In this thesis we use the following Sobolev spaces:

$$\begin{aligned} H^1(\Omega) &:= \{f \in L^2(\Omega) \mid \nabla f \in L^2(\Omega, \mathbb{R}^d)\}, \\ H(\operatorname{div}, \Omega) &:= \{v \in L^2(\Omega, \mathbb{R}^d) \mid \operatorname{div} v \in L^2(\Omega)\}, \\ H(\operatorname{curl}, \Omega) &:= \begin{cases} \{v \in L^2(\Omega, \mathbb{R}^3) \mid \operatorname{curl} v \in L^2(\Omega, \mathbb{R}^3)\} & \text{if } d = 3 \\ \{v \in L^2(\Omega, \mathbb{R}^2) \mid \operatorname{curl} v \in L^2(\Omega)\} & \text{if } d = 2 \end{cases}. \end{aligned}$$

Note that the definition of $H(\operatorname{curl}, \Omega)$ depends on the dimension. These spaces are equipped with the following inner products:

$$\begin{aligned} (f, g)_{H^1} &:= (f, g) + (\nabla f, \nabla g), \\ (v, w)_{H(\operatorname{div})} &:= (v, w) + (\operatorname{div} v, \operatorname{div} w), \\ (v, w)_{H(\operatorname{curl})} &:= (v, w) + (\operatorname{curl} v, \operatorname{curl} w), \end{aligned}$$

and the corresponding norms $\|\cdot\|_{H^1}$, $\|\cdot\|_{H(\operatorname{div})}$, and $\|\cdot\|_{H(\operatorname{curl})}$.

For convenience, we also define the space of divergence-free functions

$$H(\operatorname{div}_0, \Omega) := \{v \in H(\operatorname{div}, \Omega) \mid \operatorname{div} v = 0\}.$$

We define the spaces with homogenous boundary conditions as closures of infinitely differentiable functions:

$$\begin{aligned} C^{\infty}_{0,\Gamma} &:= \{\phi \in C^{\infty}_0(\mathbb{R}^d) \mid \operatorname{dist}(\operatorname{supp} \phi, \Gamma) > 0\}, \\ H^1_{0,\Gamma}(\Omega) &:= \overline{C^{\infty}_{0,\Gamma}(\Omega)}^{H^1(\Omega)}, && (2.7) \\ H_{0,\Gamma}(\operatorname{div}, \Omega) &:= \overline{C^{\infty}_{0,\Gamma}(\Omega)}^{H(\operatorname{div},\Omega)}, && (2.8) \\ H_{0,\Gamma}(\operatorname{curl}, \Omega) &:= \overline{C^{\infty}_{0,\Gamma}(\Omega)}^{H(\operatorname{curl},\Omega)}, && (2.9) \end{aligned}$$

where $\Gamma \subset \partial\Omega$. If the boundary condition is set on the whole boundary, we will omit the boundary segment form the subindex, i.e., $H^1_{0,\partial\Omega}(\Omega) = H^1_0(\Omega)$.

We also note that if f and v are smooth functions, and $\partial\Omega$ is smooth, we have

$$\begin{aligned} f \in H^1_{0,\Gamma}(\Omega) \quad &\Leftrightarrow \quad f = 0 \text{ on } \Gamma, \\ v \in H_{0,\Gamma}(\operatorname{div}, \Omega) \quad &\Leftrightarrow \quad v \cdot n = 0 \text{ on } \Gamma, \\ v \in H_{0,\Gamma}(\operatorname{curl}, \Omega) \quad &\Leftrightarrow \quad v \times n = 0 \text{ on } \Gamma, \end{aligned}$$

where $n = (n_1, n_2, \ldots, n_d)^T \in \mathbb{R}^d$ denotes the outward unit normal to the boundary $\partial\Omega$.

Useful identities

We have for all $v \in H_{0,\Gamma_N}(\mathrm{div}, \Omega)$ and $f \in H^1_{0,\Gamma_D}(\Omega)$

$$(\mathrm{div}\, v, f) = -(v, \nabla f). \tag{2.10}$$

Similar identities exist also for the rotation operators (see, e.g., [16]). If $d = 3$, we have for all $v \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$ and $w \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$

$$(v, \mathrm{curl}\, w) = (\mathrm{curl}\, v, w), \tag{2.11}$$

and if $d = 2$, we have for $v \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$ and $f \in H^1_{0,\Gamma_N}(\Omega)$

$$(\mathrm{curl}\, v, f) = (v, \underline{\mathrm{curl}}\, f). \tag{2.12}$$

We also note that, for smooth f, v, and $\partial\Omega$, the identity (2.10) comes from the Gauß' theorem

$$(\mathrm{div}\, v, f) = -(v, \nabla f) + (v \cdot n, f)_{\partial\Omega}.$$

Constants in embedding inequalities

First of all we recall the following inequality for scalar functions f, often called the Friedrichs inequality:

$$\|f\| \leq C_F \|\nabla f\| \qquad \forall f \in H^1_0(\Omega), \tag{2.13}$$

where $0 < C_F := \lambda_1^{-1/2} < \infty$. Here λ_1 is the first Dirichlet eigenvalue of the Laplacian. The Poincaré inequality reads as

$$\|f\| \leq C_P \|\nabla f\| \qquad \forall f \in \left\{ g \in H^1(\Omega) \;\middle|\; \int_\Omega g \,\mathrm{dx} = 0 \right\}, \tag{2.14}$$

where $0 < C_P := \mu_2^{-1/2} < \infty$. Here μ_2 is the second Neumann eigenvalue of the Laplacian. These inequalities hold for bounded Lipschitz domains by the Rellich's selection theorem. It is well known that $C_F < C_P$ holds for $d \in \{2, 3\}$ (see, e.g., [11]). In [38] it was proven that for convex domains the following estimate holds:

$$C_P \leq \frac{l_\Omega}{\pi}, \tag{2.15}$$

where l_Ω is the diameter of Ω.

For simple bounded domains, some analytically obtained upper bounds for C_F sharper than (2.15) are known. For a domain in $\mathbb{R}^d$ which is encompassed inside a rectangle with edge lengths $l_1, l_2, \ldots, l_d$, we have

$$C_F \leq \left(\frac{1}{\pi^2 (l_1^{-2} + l_2^{-2} + \cdots + l_d^{-2})} \right)^{1/2}. \tag{2.16}$$

This result can be found in [24].

These constants can also be approximated computationally. A "brute force" approach to obtain lower bounds for C_F is used for example in [46]. As the size of used meshes increase, the closer the lower bound gets to the exact value. In [17, 42] the authors expose two different ways to computationally obtain bounds of these constants.

It is easy to show that inequalities similar to (2.13) and (2.14) hold also for all vector functions v:

$$\begin{aligned} &\|v\| \leq C_F \|\nabla v\| \qquad \forall v \in H_0^1(\Omega, \mathbb{R}^d), \\ &\|v\| \leq C_P \|\nabla v\| \qquad \forall v \in \left\{ w \in H^1(\Omega, \mathbb{R}^d) \;\middle|\; \int_\Omega w_i \,\mathrm{dx} = 0,\ i \in \{1, 2, \ldots, d\} \right\}. \end{aligned}$$

A similar result holds for the operator curl in bounded domains. Assuming that $\partial\Omega$ is connected we have

$$\|v\| \leq C_M \|\operatorname{curl} v\| \qquad \forall v \in H_0(\operatorname{curl}, \Omega) \cap H(\operatorname{div}_0, \Omega), \tag{2.17}$$

where $0 < C_M < \infty$.

If $d = 2$, the constant C_M can be replaced by the constant C_P in (2.17). This is easily proven by using a so-called "stream function", a rotation of a potential function. We take $v \in H_0(\operatorname{curl}, \Omega) \cap H(\operatorname{div}_0, \Omega)$, and represent it by $v = \underline{\operatorname{curl}}\, \varphi$, where $\varphi \in H^1(\Omega)$ is such that $\int_\Omega \varphi \,\mathrm{dx} = 0$. By (2.12) and the Cauchy-Schwarz inequality (2.3) we can write

$$\|v\|^2 = (v, \underline{\operatorname{curl}}\, \varphi) = (\operatorname{curl} v, \varphi) \leq \|\operatorname{curl} v\| \|\varphi\| \leq C_P \|\operatorname{curl} v\| \|\nabla \varphi\|,$$

where in the last step we used (2.14). Since $\|v\| = \|\underline{\operatorname{curl}}\, \varphi\| = \|\nabla \varphi\|$ we obtain

$$\|v\| \leq C_P \|\operatorname{curl} v\|. \tag{2.18}$$

A recent result by D. Pauly shows that if $d = 3$ and Ω is convex, the constant C_M is bounded by the Poincaré constant, i.e., we have $C_M \leq C_P$ (see [33–35]).

Meshes and finite elements

In this thesis we will use Nédélec elements of the first family [29] to discretize our model problem. Edge based elements include also the Nédélec elements of the second family [28] and the Raviart-Thomas elements [40].

We denote by $\mathcal{T}_h$ the partition of Ω to the union of non-overlapping tetrahedras if $d = 3$, and triangles if $d = 2$. An element in $\mathcal{T}_h$ is denoted by T. For any $\mathrm{T} \in \mathcal{T}_h$, the sets $\mathcal{E}(\mathrm{T})$ and $\mathcal{N}(\mathrm{T})$ denote the edges and nodes of T, respectively. The sets

$$\mathcal{E}_h := \bigcup_{\mathrm{T} \in \mathcal{T}_h} \mathcal{E}(\mathrm{T}) \qquad \text{and} \qquad \mathcal{N}_h := \bigcup_{\mathrm{T} \in \mathcal{T}_h} \mathcal{N}(\mathrm{T})$$

contain all edges and all nodes of $\mathcal{T}_h$, respectively. The sets

$$\omega_{\mathrm{E}} := \bigcup_{\mathrm{E} \in \mathcal{E}(\mathrm{T}')} \mathrm{T}' \qquad \text{and} \qquad \omega_{\mathrm{N}} := \bigcup_{\mathrm{N} \in \mathcal{N}(\mathrm{T}')} \mathrm{T}'$$

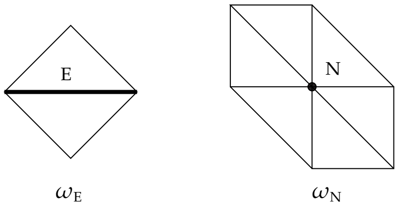


FIGURE 2 The element patches ω_{E} and ω_{N} when $d = 2$.

define patches of elements associated with a given edge $\mathrm{E} \in \mathcal{E}_h$ and node $\mathrm{N} \in \mathcal{N}_h$, respectively. These element patches are visualized in Figure 2 when $d = 2$.

For every $\mathrm{E} \in \mathcal{E}_h$, we assign a vector t_{E} of unitary length, which is tangent to E. We use $|\cdot|$ to denote area of a domain or length of an edge. The number of elements in a set is denoted by $\#(\cdot)$.

2.1 A Maxwell type problem: the eddy-current problem

In classical settings the Maxwell problem is defined by E, D (electric field and induction), H and B (magnetic field and induction) satisfying

$$\partial_t D - \operatorname{curl} H = -J, \qquad \text{Ampère's law} \qquad (2.19)$$

$$\partial_t B + \operatorname{curl} E = 0, \qquad \text{Faraday's law} \qquad (2.20)$$

$$\operatorname{div} D = \rho, \qquad \text{Gauss' law} \qquad (2.21)$$

$$\operatorname{div} B = 0, \qquad \text{Gauss' law of magnetism} \qquad (2.22)$$

for all (t, x) in $(0, T) \times \Omega$. Here the variable $t \in (0, T)$ denotes time, $T > 0$, and ∂_t is the differential with respect to time. There are two sources in these equations: J is the electric current density, and ρ is the electric charge density. In linear media the constituent relations are

$$D = \epsilon E, \qquad (2.23)$$

$$B = \mu H, \qquad (2.24)$$

where the electric permittivity ϵ and the magnetic permeability μ are bounded, uniformly positive definite, real, and symmetric. We also assume them to be time-independent. By using the constituent relations (2.23) and (2.24) we can rewrite the Maxwell equations (2.19)–(2.22) in terms of E and H only:

$$\begin{aligned} \epsilon \partial_t E - \operatorname{curl} H &= -J, \\ \mu \partial_t H + \operatorname{curl} E &= 0. \end{aligned}$$

These equations must be accompanied by initial conditions and suitable boundary conditions. By using the backward-Euler scheme for the time-derivatives, we have

$$\begin{aligned}
&\frac{\epsilon}{\triangle t}\left(E^i - E^{i-1}\right) - \operatorname{curl} H^i = -J^i,\\
&\frac{\mu}{\triangle t}\left(H^i - H^{i-1}\right) + \operatorname{curl} E^i = 0, \qquad i = 1, \dots, N,\ N = \frac{T}{\triangle t},
\end{aligned}$$

where $\triangle t$ is the size of the time step. By eliminating H^i and transferring E^{i-1} and H^{i-1} to the right hand side, we obtain

$$\operatorname{curl}\left(\mu^{-1}\operatorname{curl} E^i\right) + \frac{\epsilon}{(\triangle t)^2}E^i = \frac{1}{\triangle t}\left(-J^i + \frac{\epsilon}{\triangle t}E^{i-1} + \operatorname{curl} H^{i-1}\right).$$

We denote the right hand side by F, set $\kappa = \epsilon(\triangle t)^{-2}$, omit the superscript i, and arrive at the following coercive problem: find the electric field E such that

$$\operatorname{curl}\mu^{-1}\operatorname{curl} E + \kappa E = F \qquad \text{in } \Omega, \tag{2.25}$$

$$E \times n = 0 \qquad \text{on } \Gamma_D, \tag{2.26}$$

$$\mu^{-1}\operatorname{curl} E \times n = 0 \qquad \text{on } \Gamma_N. \tag{2.27}$$

Here we have also added mixed boundary conditions: n denotes the outward unit normal to the boundary $\partial\Omega$. The boundary condition (2.26) is of the Dirichlet type (in literature this boundary condition is often referred to as the "perfect electric conductor" boundary condition, or shortly, the PEC boundary condition). The boundary condition (2.27) is of the Neumann type. In literature the equation (2.25) is often referred to as the *eddy current* equation.

For the rest of the thesis, the problem (2.25)–(2.27) is considered as it is, without the original time-dependence, and without considering $F \in L^2(\Omega)$ to depend on other functions. We will still call μ the magnetic permeability, and we will refer to κ as the electric permittivity, since it essentially depends only on ϵ. We assume these scalar material parameters belong to $L^\infty(\overline{\Omega})$, and that

$$\begin{aligned}
0 < \underline{\mu} \leq \mu(x) \leq \overline{\mu} < \infty,\\
0 < \underline{\kappa} \leq \kappa(x) \leq \overline{\kappa} < \infty,
\end{aligned}$$

for a.e. $x \in \overline{\Omega}$.

In mixed form, the model problem (2.25)–(2.27) reads: find the electric and magnetic fields $E, H \in H(\operatorname{curl}, \Omega)$ such that

$$\operatorname{curl} H + \kappa E = F \qquad \text{in } \Omega, \tag{2.28}$$

$$H = \mu^{-1}\operatorname{curl} E \qquad \text{in } \Omega, \tag{2.29}$$

$$E \times n = 0 \qquad \text{on } \Gamma_D, \tag{2.30}$$

$$H \times n = 0 \qquad \text{on } \Gamma_N. \tag{2.31}$$

The exact solution pair (E, H) belongs then to $H_{0,\Gamma_D}(\operatorname{curl}, \Omega) \times H_{0,\Gamma_N}(\operatorname{curl}, \Omega)$.

Generalized solutions

By multiplying (2.25)–(2.27) with a test function $v \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$, and integrating over Ω, we obtain

$$(\mathrm{curl}\, \mu^{-1}\mathrm{curl}\, E, v) + (\kappa E, v) = (F, v) \qquad \forall v \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega).$$

By applying (2.11) to the first term on the left hand side, and taking into account (2.1)–(2.2), we obtain the corresponding weak formulation

$$(\mathrm{curl}\, E, \mathrm{curl}\, v)_{\mu^{-1}} + (E, v)_\kappa = (F, v) \qquad \forall v \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega). \tag{2.32}$$

We call (2.32) the *primal problem*. The *dual problem* for the dual variable, the magnetic field H, is obtained by multiplying (2.28)–(2.31) with $\kappa^{-1}\mathrm{curl}\, q$ for any $q \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$:

$$(\mathrm{curl}\, H, \kappa^{-1}\mathrm{curl}\, q) + (\kappa E, \kappa^{-1}\mathrm{curl}\, q) = (F, \kappa^{-1}\mathrm{curl}\, q) \qquad \forall q \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega).$$

By applying (2.11) to the second term on the left hand side, and noting that indeed $E \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$, we obtain the weak formulation

$$(\mathrm{curl}\, H, \mathrm{curl}\, q)_{\kappa^{-1}} + (H, q)_\mu = (F, \mathrm{curl}\, q)_{\kappa^{-1}} \qquad \forall q \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega). \tag{2.33}$$

It is well known that the solutions of the primal and dual problems exist and are unique, and moreover, coincide with the solution pair of the strong problem (see, e.g., [PIII, Section 2]). By defining

$$\begin{aligned}
a(u, w) &:= (\mathrm{curl}\, u, \mathrm{curl}\, w)_{\mu^{-1}} + (u, w)_\kappa, \\
l(w) &:= (f, w), \\
\hat{a}(u, w) &:= (\mathrm{curl}\, u, \mathrm{curl}\, w)_{\kappa^{-1}} + (u, w)_\mu, \\
\hat{l}(w) &:= (f, \mathrm{curl}\, w)_{\kappa^{-1}},
\end{aligned}$$

the problems (2.32) and (2.33) can be written concisely as

$$a(E, v) = l(v) \qquad \text{and} \qquad \hat{a}(H, q) = \hat{l}(q),$$

respectively. The bilinear forms $a(\cdot, \cdot)$ and $\hat{a}(\cdot, \cdot)$ define the energy norms

$$\begin{aligned}
|\!|\!|\, v \,|\!|\!|^2 &:= a(v, v) = \|\mathrm{curl}\, v\|^2_{\mu^{-1}} + \|v\|^2_\kappa, \\
|\!|\!|[\, q \,]|\!|\!|^2 &:= \hat{a}(q, q) = \|\mathrm{curl}\, q\|^2_{\kappa^{-1}} + \|q\|^2_\mu,
\end{aligned}$$

both of which are equivalent to the norm $\| \cdot \|_{H(\mathrm{curl})}$. The functional type a posteriori error estimates estimate the error in these type weighed norms. We also define the following combined norm

$$|[v, q]|^2 = |\!|\!|\, v \,|\!|\!|^2 + |\!|\!|[\, q \,]|\!|\!|^2$$

for the purpose of measuring the error in both primal and dual variables simultaneously.

The problems (2.32) and (2.33) can also be expressed by the minimization of the following energy functionals

$$\mathcal{J}(v) := \frac{1}{2}a(v,v) - l(v) \qquad \text{and} \qquad \hat{\mathcal{J}}(q) := \frac{1}{2}\hat{a}(q,q) - \hat{l}(q) \tag{2.34}$$

with respect to $v \in H_{0,\Gamma_D}(\mathrm{curl},\Omega)$ and $q \in H_{0,\Gamma_N}(\mathrm{curl},\Omega)$. It is well known that this minimization (see, e.g., [9]) leads to the exact solutions:

$$\min_v \mathcal{J}(v) = \mathcal{J}(E) \qquad \text{and} \qquad \min_q \hat{\mathcal{J}}(q) = \hat{\mathcal{J}}(H).$$

Remark 2.1. *By adding the weak forms* (2.32) *and* (2.33) *together, and choosing $v = E$ and $q = H$, we see that*

$$|\!|\!| E |\!|\!|^2 + [\![H]\!]^2 = (F,E) + (F, \mathrm{curl}\, H)_{\kappa^{-1}} = (F, \kappa E + \mathrm{curl}\, H)_{\kappa^{-1}},$$

and by (2.28) *we have*

$$|[E,H]|^2 = (F,F)_{\kappa^{-1}} = \|F\|^2_{\kappa^{-1}},$$

so the solution mapping $\mathcal{S} : L^2 \to H_{0,\Gamma_D}(\mathrm{curl},\Omega) \times H_{0,\Gamma_N}(\mathrm{curl},\Omega)$ is an isometry *under the combined energy norm, i.e., $|S| = 1$. This property of the solution mapping is fulfilled in all problems of this type (see [PIII, Remark 3.18]).*

Remark 2.2. *If $d = 2$, the double* curl *in* (2.25) *should be understood as* $\underline{\mathrm{curl}}\,\mathrm{curl}$*, i.e., in two dimensions, the mixed problem* (2.28)–(2.31) *would read: find the electric field $E \in H(\mathrm{curl},\Omega)$ and the magnetic field $H \in H^1(\Omega)$ such that*

$$\begin{aligned} \underline{\mathrm{curl}}\, H + \kappa E &= F && \textit{in } \Omega, \\ H &= \mu^{-1}\mathrm{curl}\, E && \textit{in } \Omega, \\ E \times n &= 0 && \textit{on } \Gamma_D, \\ H &= 0 && \textit{on } \Gamma_N. \end{aligned}$$

The exact solution pair (E,H) belongs then to $H_{0,\Gamma_D}(\mathrm{curl},\Omega) \times H^1_{0,\Gamma_N}(\Omega)$.

3 FUNCTIONAL A POSTERIORI ERROR CONTROL

Research on a posteriori error control for Maxwell type problems has been mainly done in the context of the residual approach. Residual type error estimates and indicators were studied in [5, 25, 26], and an equilibrated residual approach was presented in [7]. A posteriori estimates for non-conforming approximations were studied in [14]. A Zienkiewicz-Zhu type error estimate was introduced in [32].

In this thesis we will concentrate in the functional type a posteriori error estimates developed by S. Repin. A consequent exposition of the corresponding theory is given in the books [22, 31, 41]. The functional type estimates do not rely on any properties of the numerical method used to compute approximate solutions. This means that a posteriori estimates of the functional type are valid for any *conforming* approximation. A conforming approximation of the electric field E of the system (2.25)–(2.27) would be a function which belongs to $H(\mathrm{curl}, \Omega)$ and satisfies the Dirichlet boundary condition (2.26) exactly. Another important property of these estimates is that they do not contain mesh dependent constants. The constants appearing in these estimates are global constants arising from embedding inequalities. In this thesis, the presented estimates either contain no constants, or contain the constants from (2.13) and (2.17).

Functional type estimates for conforming approximations of the eddy current problem (2.25)–(2.26) with $\Gamma_D = \partial\Omega$ were originally derived in [30, 43] by S. Repin and P. Neittaanmäki, and in [12] by A. Hannukainen. Functional type estimates for approximations of the static and time-dependent Maxwell problem were derived in [36] by D. Pauly and S. Repin, and in [37] by D. Pauly, S. Repin, and T. Rossi.

In this thesis we concentrate only in the case of conforming approximations for the eddy-current problem. In the following sections we present all essential results citing the original works where they were derived in.

3.1 Classical lower bounds

In literature, the functional type lower bounds are often called *minorants*, and they answer the question "how much error the approximation contains at least?"

Theorem 3.1. *Let* $E \in H_{0,\Gamma_D}(\text{curl}, \Omega)$ *be the exact electric field of the system* (2.28)–(2.31). *Then, for an arbitrary* $\widetilde{E} \in H_{0,\Gamma_D}(\text{curl}, \Omega)$ *we have*

$$|\!|\!| E - \widetilde{E} |\!|\!|^2 \geq m_E(\widetilde{E}, Z) \qquad \forall Z \in H_{0,\Gamma_D}(\text{curl}, \Omega),$$

where

$$m_E(\widetilde{E}, Z) := 2(\mathcal{J}(\widetilde{E}) - \mathcal{J}(Z)).$$

Proof. We have

$$\begin{aligned} |\!|\!| E - \widetilde{E} |\!|\!|^2 &= a(E - \widetilde{E}, E - \widetilde{E}) \\ &= a(E, E) + a(\widetilde{E}, \widetilde{E}) - 2a(E, \widetilde{E}) \\ &= a(E, E) + a(\widetilde{E}, \widetilde{E}) - 2l(\widetilde{E}). \end{aligned}$$

By adding $2(-a(E, E) + l(E)) = 0$ to the right hand side, we obtain

$$\begin{aligned} |\!|\!| E - \widetilde{E} |\!|\!|^2 &= a(\widetilde{E}, \widetilde{E}) - 2l(\widetilde{E}) - a(E, E) + 2l(E) \\ &= 2(\mathcal{J}(\widetilde{E}) - \mathcal{J}(E)). \end{aligned}$$

Since by definition $J(E) \leq J(Z)$ for any $Z \in H_{0,\Gamma_D}(\text{curl}, \Omega)$, we obtain the result.
□

Theorem 3.2. *Let* $H \in H_{0,\Gamma_N}(\text{curl}, \Omega)$ *be the exact magnetic field of the system* (2.28)–(2.31). *Then, for an arbitrary* $\widetilde{H} \in H_{0,\Gamma_N}(\text{curl}, \Omega)$ *we have*

$$[\![H - \widetilde{H}]\!]^2 \geq m_H(\widetilde{H}, \hat{Z}) \qquad \forall \hat{Z} \in H_{0,\Gamma_N}(\text{curl}, \Omega),$$

where

$$m_H(\widetilde{H}, \hat{Z}) := 2(\hat{\mathcal{J}}(\widetilde{H}) - \hat{\mathcal{J}}(\hat{Z})).$$

Proof. The proof is identical to the proof of Theorem 3.1.
□

Both of the minorants are *sharp*, i.e., they do not contain a gap between the exact errors and the estimates:

$$\begin{aligned} \max_{Z} m_E(\widetilde{E}, Z) &= m_E(\widetilde{E}, E) = |\!|\!| E - \widetilde{E} |\!|\!|^2, \\ \max_{\hat{Z}} m_H(\widetilde{H}, \hat{Z}) &= m_H(\widetilde{H}, H) = [\![H - \widetilde{H}]\!]^2 . \end{aligned}$$

Finally, we note that the functional a posteriori error minorants m_E and m_H are both fully computable: they contain only the problem data, conforming numerical approximations $\widetilde{E}$ and $\widetilde{H}$, and the arbitrary functions Z and $\hat{Z}$.

A lower bound of the error in the primal variable was originally derived in [43, Chapter 2]. This lower bound can also be found in [30, Section 3.2], the included paper [PI, Proposition 2.2], and the book [22, Section 4.3]. In all of these papers the lower bound is derived in a different way than in Theorem 3.1, but they are nevertheless equivalent.

3.2 Classical upper bounds

In a posteriori error estimation one is often more interested in upper bounds of errors, in order to obtain information of the feasibility of an approximation. An upper bound, or a *majorant*, answers the question "how much error the approximation contains at most?" The derivation of functional type a posteriori error estimates directly utilize the weak formulations of the problems at hand.

Theorem 3.3. *Let* $E \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$ *be the exact electric field of the system* (2.28)–(2.31). *Then, for an arbitrary* $\widetilde{E} \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$ *we have*

$$|\!|\!| E - \widetilde{E} |\!|\!|^2 \leq \mathcal{M}_E(\widetilde{E}, Y) \qquad \forall Y \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega),$$

where

$$\mathcal{M}_E(\widetilde{E}, Y) := \|F - \kappa\widetilde{E} - \mathrm{curl}\, Y\|^2_{\kappa^{-1}} + \|Y - \mu^{-1}\mathrm{curl}\, \widetilde{E}\|^2_{\mu}. \tag{3.1}$$

Proof. First, we add $(\mathrm{curl}\, \widetilde{E}, \mathrm{curl}\, v)_{\mu^{-1}} + (\widetilde{E}, v)_\kappa$ to the both sides of the weak form (2.32), and obtain

$$(\mathrm{curl}(E - \widetilde{E}), \mathrm{curl}\, v)_{\mu^{-1}} + (E - \widetilde{E}, v)_\kappa = (F - \kappa\widetilde{E}, v) - (\mu^{-1}\mathrm{curl}\, \widetilde{E}, \mathrm{curl}\, v). \tag{3.2}$$

By (2.11) we see that for an arbitrary function $Y \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$ and any function $v \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$, we have $(Y, \mathrm{curl}\, v) - (\mathrm{curl}\, Y, v) = 0$. By adding this to the right hand side of (3.2) it becomes

$$\begin{aligned} & (F - \kappa\widetilde{E} - \mathrm{curl}\, Y, v) + (Y - \mu^{-1}\mathrm{curl}\, \widetilde{E}, \mathrm{curl}\, v) && (3.3)\\ &= (\kappa^{-1/2}[F - \kappa\widetilde{E} - \mathrm{curl}\, Y], \kappa^{1/2}v) + (\mu^{1/2}[Y - \mu^{-1}\mathrm{curl}\, \widetilde{E}], \mu^{-1/2}\mathrm{curl}\, v) \\ &\overset{(2.3)}{\leq} \|F - \kappa\widetilde{E} - \mathrm{curl}\, Y\|_{\kappa^{-1}}\|v\|_\kappa + \|Y - \mu^{-1}\mathrm{curl}\, \widetilde{E}\|_\mu \|\mathrm{curl}\, v\|_{\mu^{-1}} \\ &\overset{(2.4)}{\leq} \left(\|F - \kappa\widetilde{E} - \mathrm{curl}\, Y\|^2_{\kappa^{-1}} + \|Y - \mu^{-1}\mathrm{curl}\, \widetilde{E}\|^2_\mu\right)^{1/2} |\!|\!| v |\!|\!| \,. && (3.4) \end{aligned}$$

By choosing $v = E - \widetilde{E} \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega)$ in (3.2) and (3.4) (and squaring both sides) we obtain the estimate

$$|\!|\!| E - \widetilde{E} |\!|\!|^2 \leq \|F - \kappa\widetilde{E} - \mathrm{curl}\, Y\|^2_{\kappa^{-1}} + \|Y - \mu^{-1}\mathrm{curl}\, \widetilde{E}\|^2_\mu = \mathcal{M}_E(\widetilde{E}, Y)$$

for an arbitrary Y.
□

It is easy to see that this majorant is sharp:

$$\min_Y \mathcal{M}_E(\widetilde{E}, Y) = \mathcal{M}_E(\widetilde{E}, H) = |\!|\!| E - \widetilde{E} |\!|\!|^2 \,.$$

This immediately shows us that (3.1) provides us the means to obtain approximations for the exact magnetic field H. By globally minimizing $\mathcal{M}_E(\widetilde{E}, Y)$ with respect to Y we obtain the following problem: Find $Y \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$ such that

$$(\mathrm{curl}\, Y, \mathrm{curl}\, q)_{\kappa^{-1}} + (Y, q)_\mu = (F, \mathrm{curl}\, q)_{\kappa^{-1}} \qquad \forall q \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega).$$

Indeed, this problem is exactly the dual problem (2.33). Global minimization of functional majorants is discussed in detail in Subsection 3.5.1.

Using the same arguments as in proving Theorem 3.3, we can derive an upper bound for the dual problem (2.33).

Theorem 3.4. *Let $H \in H_{0,\Gamma_N}(\operatorname{curl}, \Omega)$ be the exact magnetic field of the system* (2.28)–(2.31). *Then, for an arbitrary $\widetilde{H} \in H_{0,\Gamma_N}(\operatorname{curl}, \Omega)$ we have*

$$|\!|\!|\, H - \widetilde{H} \,|\!|\!|^2 \leq \mathcal{M}_H(\widetilde{H}, X) \qquad \forall X \in H_{0,\Gamma_D}(\operatorname{curl}, \Omega),$$

where

$$\mathcal{M}_H(\widetilde{H}, X) := \|F - \kappa X - \operatorname{curl} \widetilde{H}\|^2_{\kappa^{-1}} + \|\widetilde{H} - \mu^{-1} \operatorname{curl} X\|^2_{\mu}. \tag{3.5}$$

Proof. The proof is nearly identical to the proof of Theorem 3.3. First, we add $(\operatorname{curl} \widetilde{H}, \operatorname{curl} q)_{\kappa^{-1}} + (\widetilde{H}, q)_\mu$ to the both sides of the weak form (2.33), and obtain

$$(\operatorname{curl}(H - \widetilde{H}), \operatorname{curl} q)_{\kappa^{-1}} + (H - \widetilde{H}, q)_\mu = (F - \operatorname{curl} \widetilde{H}, \operatorname{curl} q)_{\kappa^{-1}} - (\widetilde{H}, q)_\mu. \tag{3.6}$$

By (2.11) we see that for any $q \in H_{0,\Gamma_N}(\operatorname{curl}, \Omega)$ and $X \in H_{0,\Gamma_D}(\operatorname{curl}, \Omega)$, we have $(q, \operatorname{curl} X) - (\operatorname{curl} q, X) = 0$. By adding this to the right hand side of (3.6) it becomes

$$\begin{aligned}
& (F - \kappa X - \operatorname{curl} \widetilde{H}, \operatorname{curl} q)_{\kappa^{-1}} - (\widetilde{H} - \mu^{-1} \operatorname{curl} X, q)_\mu && (3.7) \\
\overset{(2.3)}{\leq}\; & \|F - \kappa X - \operatorname{curl} \widetilde{H}\|_{\kappa^{-1}} \|\operatorname{curl} q\|_{\kappa^{-1}} + \|\widetilde{H} - \mu^{-1} \operatorname{curl} X\|_\mu \|q\|_\mu && \\
\overset{(2.4)}{\leq}\; & \left(\|F - \kappa X - \operatorname{curl} \widetilde{H}\|^2_{\kappa^{-1}} + \|\widetilde{H} - \mu^{-1} \operatorname{curl} X\|^2_\mu \right)^{1/2} |\!|\!|\, q \,|\!|\!| \,. && (3.8)
\end{aligned}$$

By choosing $q = H - \widetilde{H} \in H_{0,\Gamma_N}(\operatorname{curl}, \Omega)$ in (3.6) and (3.8) (and squaring both sides) we obtain the estimate

$$|\!|\!|\, H - \widetilde{H} \,|\!|\!|^2 \leq \; \|F - \kappa X - \operatorname{curl} \widetilde{H}\|^2_{\kappa^{-1}} + \|\widetilde{H} - \mu^{-1} \operatorname{curl} X\|^2_\mu \; = \mathcal{M}_H(\widetilde{H}, X)$$

for an arbitrary X.
□

It is easy to see that the majorant (3.5) is sharp:

$$\min_X \mathcal{M}_H(\widetilde{H}, X) = \mathcal{M}_H(\widetilde{H}, E) = |\!|\!|\, H - \widetilde{H} \,|\!|\!|^2 \,.$$

As before, this means that (3.5) provides us the means to obtain approximations to the electric field E. In fact, global minimization of $\mathcal{M}_H(\widetilde{H}, X)$ with respect to X would lead to the weak form (2.32). Finally, we note that the functional a posteriori error majorants $\mathcal{M}_E$ and $\mathcal{M}_H$ are both fully computable: they contain only the problem data, conforming numerical approximations $\widetilde{E}$ and $\widetilde{H}$, and the arbitrary functions Y and X.

The upper bound for the primal variable in Theorem 3.3 was originally derived in [43, Chapter 2] and [12, Theorem 1]. This upper bound can be found also in [30, Proposition 1]. The upper bound of the error in the dual variable can be found in a more general form in [31, Section 7.2].

3.3 An advanced upper bound

The upper bound $\mathcal{M}_E$ from Theorem 3.3 is sensitive with respect to small values of κ (or, more precisely, if κ obtains considerably smaller values than μ). With some additional assumptions there is a way to overcome this effect.

For this Section we will assume there is no Neumann boundary condition. Our model problem is then the following system:

$$\operatorname{curl} \mu^{-1} \operatorname{curl} E + \kappa E = F \qquad \text{in } \Omega, \tag{3.9}$$

$$E \times n = 0 \qquad \text{on } \partial\Omega, \tag{3.10}$$

where the magnetic permeability $\mu \in L^\infty(\overline{\Omega})$. By assuming that the electric permittivity κ is a constant, and that the source $F \in H(\operatorname{div}_0, \Omega)$, we see (by taking the divergence of (3.9)) that $\operatorname{div} F = \operatorname{div} \kappa E = 0$, so $\operatorname{div} E = 0$. The weak formulation for this problem is derived in the same way as for the more general problem (2.25)–(2.27), and it is

$$(\operatorname{curl} E, \operatorname{curl} v)_{\mu^{-1}} + (E, v)_\kappa = (F, v) \qquad \forall v \in H_0(\operatorname{curl}, \Omega). \tag{3.11}$$

In the following proof we need that the approximation $\widetilde{E}$ has global divergence, i.e., it needs to belong to $H_0(\operatorname{curl}, \Omega) \cap H(\operatorname{div}, \Omega)$. In proving the advanced form of the majorant the inequality (2.17) is used, so we also need to assume that $\partial\Omega$ is connected.

Theorem 3.5. *Let $\partial\Omega$ be connected, and $E \in H(\operatorname{curl}, \Omega) \cap H(\operatorname{div}_0, \Omega)$ be the exact solution of the system* (3.9)–(3.10). *Then, for an arbitrary $\widetilde{E} \in H_0(\operatorname{curl}, \Omega) \cap H(\operatorname{div}, \Omega)$ we have*

$$|\!|\!| E - \widetilde{E} |\!|\!|^2 \leq \mathcal{M}_E^{(\lambda)}(\widetilde{E}, Y) \qquad \forall Y \in H(\operatorname{curl}, \Omega),$$

where

$$\mathcal{M}_E^{(\lambda)}(\widetilde{E}, Y) := 2C_F \|(1-\lambda) r(\widetilde{E}, Y)\| \, \|\operatorname{div} \widetilde{E}\| + \|\lambda r(\widetilde{E}, Y)\|^2_{\kappa^{-1}} + \\ + \left(C_M \overline{\mu}^{1/2} \|(1-\lambda) r(\widetilde{E}, Y)\| + \|d(\widetilde{E}, Y)\|_\mu \right)^2, \tag{3.12}$$

and

$$r(\widetilde{E}, Y) := F - \kappa \widetilde{E} - \operatorname{curl} Y, \tag{3.13}$$

$$d(\widetilde{E}, Y) := Y - \mu^{-1} \operatorname{curl} \widetilde{E}. \tag{3.14}$$

Here $\lambda \in [0, 1]$.

Proof. See [PI, Proposition 2.1 and Corollary 2.1] for the proof in the case $d = 2$.[1] The proof in the case of $d = 3$ is almost identical, but we present it here for the

[1] We note that in [PI, equation (2.2)] we incorrectly use the constant C_F (denoted in this paper by C_Ω). This constant should be C_M, which in the case $d = 2$ can be replaced by C_P, as stated in Remark 3.2.

convenience of the reader. First, we add $(\operatorname{curl}\widetilde{E}, \operatorname{curl} v)_{\mu^{-1}} + (\widetilde{E}, v)_\kappa$ to the both sides of the weak form (3.11), and obtain

$$(\operatorname{curl}(E-\widetilde{E}), \operatorname{curl} v)_{\mu^{-1}} + (E-\widetilde{E}, v)_\kappa = (F - \kappa\widetilde{E}, v) - (\mu^{-1}\operatorname{curl}\widetilde{E}, \operatorname{curl} v). \quad (3.15)$$

By (2.11) we see that for an arbitrary function $Y \in H(\operatorname{curl}, \Omega)$ and any function $v \in H_0(\operatorname{curl}, \Omega)$, we have $(Y, \operatorname{curl} v) - (\operatorname{curl} Y, v) = 0$. By adding this to the right hand side of (3.15) it becomes

$$\begin{aligned}
&(\operatorname{curl}(E-\widetilde{E}), \operatorname{curl} v)_{\mu^{-1}} + (E-\widetilde{E}, v)_\kappa = \\
&= (F - \kappa\widetilde{E} - \operatorname{curl} Y, v) + (Y - \mu^{-1}\operatorname{curl}\widetilde{E}, \operatorname{curl} v) \\
&= (\lambda r(\widetilde{E}, Y), v) + ((\lambda - 1) r(\widetilde{E}, Y), v) + (d(\widetilde{E}, Y), \operatorname{curl} v), \qquad (3.16)
\end{aligned}$$

where we have used (3.13)–(3.14) and $\lambda \in [0,1]$. We estimate the first term of (3.16) by

$$\begin{aligned}
(\lambda r(\widetilde{E}, Y), v) &= (\lambda \kappa^{-1/2} r(\widetilde{E}, Y), \kappa^{1/2} v) \\
&\stackrel{(2.3)}{\leq} \|\lambda r(\widetilde{E}, Y)\|_{\kappa^{-1}} \|v\|_\kappa \qquad (3.17)
\end{aligned}$$

Since v belongs to $H_0(\operatorname{curl}, \Omega)$, its Helmholtz decomposition is $v = v_0 + \nabla\phi$ where $v_0 \in H(\operatorname{div}_0, \Omega)$ and $\phi \in H^1_0(\Omega)$. Using (2.10) we see that ϕ satisfies $(\nabla\phi, \nabla\varphi) = -(\operatorname{div}\nabla\phi, \varphi) = -(\operatorname{div} v, \varphi)$ for all $\varphi \in H^1_0(\Omega)$, which implies the estimate $\|\nabla\phi\| \leq C_F \|\operatorname{div} v\|$, where C_F is the constant from the Friedrichs inequality (2.13). By using this, the second term of (3.16) can be estimated by

$$\begin{aligned}
&((\lambda - 1) r(\widetilde{E}, Y), v) = \\
&= ((\lambda - 1) r(\widetilde{E}, Y), v_0) + ((\lambda - 1) r(\widetilde{E}, Y), \nabla\phi) \\
&\stackrel{(2.3)}{\leq} \|(\lambda - 1) r(\widetilde{E}, Y)\| \, \|v_0\| + \|(\lambda - 1) r(\widetilde{E}, Y)\| \, \|\nabla\phi\| \\
&\stackrel{(2.17)}{\leq} C_M \overline{\mu}^{1/2} \|(\lambda - 1) r(\widetilde{E}, Y)\| \, \|\operatorname{curl} v\|_{\mu^{-1}} + C_F \|(\lambda - 1) r(\widetilde{E}, Y)\| \, \|\operatorname{div} v\| \qquad (3.18)
\end{aligned}$$

where we have also noted that $\operatorname{curl} v_0 = \operatorname{curl} v$. The third term in (3.16) can be estimated by

$$\begin{aligned}
(d(\widetilde{E}, Y), \operatorname{curl} v) &= (\mu^{1/2} d(\widetilde{E}, Y), \mu^{-1/2} \operatorname{curl} v) \\
&\stackrel{(2.3)}{\leq} \|d(\widetilde{E}, Y)\|_\mu \|\operatorname{curl} v\|_{\mu^{-1}} \qquad (3.19)
\end{aligned}$$

By choosing $v = E - \widetilde{E} \in H_0(\operatorname{curl}, \Omega)$ in (3.16)–(3.19) we have

$$\begin{aligned}
|\!|\!| E - \widetilde{E} |\!|\!|^2 &\leq \left(C_M \overline{\mu}^{1/2} \|(\lambda - 1) r(\widetilde{E}, Y)\| + \|d(\widetilde{E}, Y)\|_\mu \right) \|\operatorname{curl}(E - \widetilde{E})\|_{\mu^{-1}} + \\
&\quad + \|\lambda r(\widetilde{E}, Y)\|_{\kappa^{-1}} \|E - \widetilde{E}\|_\kappa + C_F \|(\lambda - 1) r(\widetilde{E}, Y)\| \, \|\operatorname{div} \widetilde{E}\| \\
&\stackrel{(2.6)}{\leq} \frac{1}{2} \left(C_M \overline{\mu}^{1/2} \|(\lambda - 1) r(\widetilde{E}, Y)\| + \|d(\widetilde{E}, Y)\|_\mu \right)^2 + \frac{1}{2} \|\operatorname{curl}(E - \widetilde{E})\|^2_{\mu^{-1}} + \\
&\quad + \frac{1}{2} \|\lambda r(\widetilde{E}, Y)\|^2_{\kappa^{-1}} + \frac{1}{2} \|E - \widetilde{E}\|^2_\kappa + C_F \|(\lambda - 1) r(\widetilde{E}, Y)\| \, \|\operatorname{div} \widetilde{E}\|.
\end{aligned}$$

Transferring the terms including the exact solution E to the left hand side, and by multiplying the inequality by 2, we obtain

$$
\begin{aligned}
|\!|\!| E - \widetilde{E} |\!|\!|^2 \leq \Big(C_M \overline{\mu}^{1/2} \| (\lambda - 1) r(\widetilde{E}, Y) \| + \| d(\widetilde{E}, Y) \|_\mu \Big)^2 + \\
+ \| \lambda r(\widetilde{E}, Y) \|^2_{\kappa^{-1}} + 2 C_F \| (\lambda - 1) r(\widetilde{E}, Y) \| \, \| \operatorname{div} \widetilde{E} \| = \mathcal{M}_E^{(\lambda)}(\widetilde{E}, Y).
\end{aligned}
$$

□

It is easy to see that the majorant (3.12) is sharp:

$$
\min_{Y,\lambda} \mathcal{M}_E^{(\lambda)}(\widetilde{E}, Y) = \mathcal{M}_E^{(1)}(\widetilde{E}, H) = |\!|\!| E - \widetilde{E} |\!|\!|^2 .
$$

Remark 3.1. *By choosing $\lambda = 1$, we obtain the classical majorant of Theorem 3.3*

$$
\mathcal{M}_E^{(1)}(\widetilde{E}, Y) := \| r(\widetilde{E}, Y) \|^2_{\kappa^{-1}} + \| d(\widetilde{E}, Y) \|^2_\mu = \mathcal{M}_E(\widetilde{E}, Y),
$$

which is suited best for large values of κ. Using this majorant with small values of κ will lead to considerable over-estimation of the error $|\!|\!| E - \widetilde{E} |\!|\!|$. However, as was shown before, this majorant is sharp. We also note, that the majorant of Theorem 3.3 is more general compared to $\mathcal{M}_E^{(1)}(\widetilde{E}, Y)$, since it does not require additional regularity of the approximation $\widetilde{E}$, or the connectedness of $\partial\Omega$. On the other hand, by choosing $\lambda = 0$, we obtain the majorant

$$
\mathcal{M}_E^{(0)}(\widetilde{E}, Y) := 2 C_F \| r(\widetilde{E}, Y) \| \, \| \operatorname{div} \widetilde{E} \| + \Big(C_M \overline{\mu}^{1/2} \| r(\widetilde{E}, Y) \| + \| d(\widetilde{E}, Y) \|_\mu \Big)^2 ,
$$

which is suited best for small values of κ. However, using this majorant with big values of κ will lead to considerable over-estimation of $|\!|\!| E - \widetilde{E} |\!|\!|$. Also, sharpness of this majorant can unfortunately not be proven.

Remark 3.2. *As stated in Chapter 2, if $d = 3$, and we additionally assume that the domain is convex, the constant C_M can be estimated from above by C_P. On the other hand, if $d = 2$, the inequality* (2.18) *can be used instead of* (2.17)*. In practice this means that the constant C_M appearing in the error majorant of Theorem 3.5 is simply replaced by C_P. However, in order to use the calculable upper bound* (2.15) *for C_P we need to assume also in this case that Ω is convex.*

Finally, we note that the functional a posteriori error majorant $\mathcal{M}_E^{(\lambda)}$ is fully computable: it contains only the problem data, a conforming numerical approximation $\widetilde{E}$, and the arbitrary function Y.

The contents of this Section is based on the included paper [PI] where Theorem 3.5 was derived for $d = 2$ and also numerically tested. However, the upper bound of Theorem 3.5 was originally derived in [30, Proposition 4], and can also be found in [22, Section 4.3].

3.4 An error equality for mixed approximations

In this section, we will understand the pair

$$(\widetilde{E},\widetilde{H}) \in H_{0,\Gamma_D}(\mathrm{curl},\Omega) \times H_{0,\Gamma_N}(\mathrm{curl},\Omega)$$

to be an conforming approximation of the exact solution

$$(E,H) \in H_{0,\Gamma_D}(\mathrm{curl},\Omega) \times H_{0,\Gamma_N}(\mathrm{curl},\Omega)$$

of the system (2.28)–(2.31). We have the following error equality:

Theorem 3.6. *For an arbitrary pair* $(\widetilde{E},\widetilde{H}) \in H_{0,\Gamma_D}(\mathrm{curl},\Omega) \times H_{0,\Gamma_N}(\mathrm{curl},\Omega)$ *we have*

$$|[(E,H)-(\widetilde{E},\widetilde{H})]|^2 = \mathcal{M}(\widetilde{E},\widetilde{H}), \tag{3.20}$$

and its normalized counterpart

$$\frac{|[(E,H)-(\widetilde{E},\widetilde{H})]|^2}{|[E,H]|^2} = \frac{\mathcal{M}(\widetilde{E},\widetilde{H})}{\|F\|^2_{\kappa^{-1}}}, \tag{3.21}$$

where

$$\mathcal{M}(\widetilde{E},\widetilde{H}) := \|F-\kappa\widetilde{E}-\mathrm{curl}\,\widetilde{H}\|^2_{\kappa^{-1}} + \|\widetilde{H}-\mu^{-1}\mathrm{curl}\,\widetilde{E}\|^2_{\mu}. \tag{3.22}$$

Proof. As the first step, we simply use (2.28) on the first term of (3.22), and insert zero in the form of $H-\mu^{-1}\mathrm{curl}E=0$ into the second term:

$$\begin{aligned}
\mathcal{M}(\widetilde{E},\widetilde{H}) &= \|F-\kappa\widetilde{E}-\mathrm{curl}\widetilde{H}\|^2_{\kappa^{-1}} + \|\widetilde{H}-\mu^{-1}\mathrm{curl}\widetilde{E}\|^2_{\mu} \\
&= \|\kappa E-\kappa\widetilde{E}+\mathrm{curl}H-\mathrm{curl}\widetilde{H}\|^2_{\kappa^{-1}} + \|\widetilde{H}-H+\mu^{-1}\mathrm{curl}E-\mu^{-1}\mathrm{curl}\widetilde{E}\|^2_{\mu} \\
&= \|E-\widetilde{E}\|^2_{\kappa} + \|\mathrm{curl}(H-\widetilde{H})\|^2_{\kappa^{-1}} + 2(\kappa(E-\widetilde{E}),\mathrm{curl}(H-\widetilde{H}))_{\kappa^{-1}} + \\
&\quad + \|\widetilde{H}-H\|^2_{\mu} + \|\mathrm{curl}(E-\widetilde{E})\|^2_{\mu^{-1}} + 2(\widetilde{H}-H,\mu^{-1}\mathrm{curl}(E-\widetilde{E}))_{\mu} \\
&= \ |\!|\!|\, E-\widetilde{E}\,|\!|\!|^2 + |\!|\!|[\, H-\widetilde{H}\,]\!|\!|^2 \\
&\quad + 2\big[(E-\widetilde{E},\mathrm{curl}(H-\widetilde{H})) + (\widetilde{H}-H,\mathrm{curl}(E-\widetilde{E}))\big].
\end{aligned} \tag{3.23}$$

Due to (2.11) the term inside the square bracets is zero. The remaining terms simply constitute the exact error in the combined norm, and we have proven the main result (3.20). Then, by Remark 2.1, or by setting $\widetilde{E}=\widetilde{H}=0$ in the equality (3.20), we have (3.21).
□

Remark 3.3. *The result of Theorem 3.6 holds also for non-homogenous boundary conditions, since due to the conformity of the approximation* $(\widetilde{E},\widetilde{H})$ *the term inside square bracts in* (3.23) *is zero. An equality can also be obtained for Robin type boundary conditions (see [PIII, Section 4] for details).*

Remark 3.4. *Note the similarity of the error majorants in Theorems 3.3, 3.4 and 3.6: by setting $Y = \widetilde{H}$ in* (3.1) *and $X = \widetilde{E}$ in* (3.5)*, we see that*

$$\mathcal{M}(\widetilde{E},\widetilde{H}) \;=\; \mathcal{M}_E(\widetilde{E},\widetilde{H}) \;=\; \mathcal{M}_H(\widetilde{H},\widetilde{E}).$$

In fact, the error equality (3.20) *directly leads to the error estimates of Theorems 3.3 and 3.4.*

Remark 3.5. *Theorem 3.6 can also be deduced as a special case of the equation [31, (7.2.14)] in the book of Neittaamäki and Repin. For the convenience of the reader we present this proof also here. We have (see the proof of Theorem 3.1)*

$$\begin{aligned} |\!|\!| E - \widetilde{E} |\!|\!|^2 \;&=\; 2(\mathcal{J}(\widetilde{E}) - \mathcal{J}(E)), \\ [\![\![H - \widetilde{H}]\!]\!]^2 \;&=\; 2(\hat{\mathcal{J}}(\widetilde{H}) - \hat{\mathcal{J}}(H)). \end{aligned}$$

By summing these two equations together, we obtain

$$|[(E,H) - (\widetilde{E},\widetilde{H})]|^2 = \;2(\mathcal{J}(\widetilde{E}) + \hat{\mathcal{J}}(\widetilde{H})) - 2(\mathcal{J}(E) + \hat{\mathcal{J}}(H)).$$

Since

$$\mathcal{J}(E) = -\frac{1}{2}\; |\!|\!| E |\!|\!|^2 \qquad \textit{and} \qquad \hat{\mathcal{J}}(H) = -\frac{1}{2}\; [\![\![H]\!]\!]^2,$$

we obtain by Remark 2.1 the following:

$$|[(E,H) - (\widetilde{E},\widetilde{H})]|^2 = \;2(\mathcal{J}(\widetilde{E}) + \hat{\mathcal{J}}(\widetilde{H})) + \|F\|^2_{\kappa^{-1}} =: \mathcal{M}_2(\widetilde{E},\widetilde{H}).$$

It is clear that $\mathcal{M}_2(\widetilde{E},\widetilde{H}) = \mathcal{M}(\widetilde{E},\widetilde{H})$. We could say that $\mathcal{M}_2$ and $\mathcal{M}$ can be regarded as positive energy functionals whose minimization will result in the exact solution (E,H), and at this point the energy will be zero.

Remark 3.6. *Note that the equality can also be shown by using the classical way of deriving functional majorants. By* (3.2) *and* (3.3) *from the proof of Theorem 3.3 we obtain*

$$\begin{aligned} &(\operatorname{curl}(E-\widetilde{E}), \operatorname{curl} v)_{\mu^{-1}} + (E-\widetilde{E}, v)_\kappa = \\ &\qquad (F - \kappa\widetilde{E} - \operatorname{curl} Y, \kappa v)_{\kappa^{-1}} + (Y - \mu^{-1}\operatorname{curl}\widetilde{E}, \mu^{-1}\operatorname{curl} v)_\mu \end{aligned} \tag{3.24}$$

for any $v \in H_{0,\Gamma_D}(\operatorname{curl},\Omega)$. On the other hand, by (3.6) *and* (3.7) *from the proof of Theorem 3.4 we obtain*

$$\begin{aligned} &(\operatorname{curl}(H-\widetilde{H}), \operatorname{curl} q)_{\kappa^{-1}} + (H-\widetilde{H}, q)_\mu = \\ &\qquad (F - \kappa X - \operatorname{curl}\widetilde{H}, \operatorname{curl} q)_{\kappa^{-1}} - (\widetilde{H} - \mu^{-1}\operatorname{curl} X, q)_\mu \end{aligned} \tag{3.25}$$

for any $q \in H_{0,\Gamma_N}(\operatorname{curl},\Omega)$. By setting $X = \widetilde{E}$ and $Y = \widetilde{H}$ and taking the sum of (3.24) *and* (3.25) *we obtain*

$$\begin{aligned} &(\operatorname{curl}(E-\widetilde{E}), \operatorname{curl} v)_{\mu^{-1}} + (E-\widetilde{E}, v)_\kappa + (\operatorname{curl}(H-\widetilde{H}), \operatorname{curl} q)_{\kappa^{-1}} + (H-\widetilde{H}, q)_\mu = \\ &\qquad (F - \kappa\widetilde{E} - \operatorname{curl}\widetilde{H}, \operatorname{curl} q + \kappa v)_{\kappa^{-1}} + (\widetilde{H} - \mu^{-1}\operatorname{curl}\widetilde{E}, \mu^{-1}\operatorname{curl} v - q)_\mu. \end{aligned} \tag{3.26}$$

By choosing $v = E - \widetilde{E}$ and $q = H - \widetilde{H}$, the left hand side of (3.26) *becomes the combined norm or the error of the approximation $(\widetilde{E}, \widetilde{H})$. Since we have*

$$\begin{aligned} \operatorname{curl} q + \kappa v &= \operatorname{curl} H - \operatorname{curl} \widetilde{H} + \kappa E - \kappa \widetilde{E} = F - \kappa \widetilde{E} - \operatorname{curl} \widetilde{H}, \\ \mu^{-1} \operatorname{curl} v - q &= \mu^{-1} \operatorname{curl} E - \mu^{-1} \operatorname{curl} \widetilde{E} - H + \widetilde{H} = \widetilde{H} - \mu^{-1} \operatorname{curl} \widetilde{E}, \end{aligned}$$

the equation (3.26) *becomes*

$$|[(E, H) - (\widetilde{E}, \widetilde{H})]|^2 = \|F - \kappa \widetilde{E} - \operatorname{curl} \widetilde{H}\|^2_{\kappa^{-1}} + \|\widetilde{H} - \mu^{-1} \operatorname{curl} \widetilde{E}\|^2_{\mu} = \mathcal{M}(\widetilde{E}, \widetilde{H}).$$

Finally, we note that the functional a posteriori error majorant $\mathcal{M}$ is fully computable: it contains only the problem data and a conforming numerical approximation $(\widetilde{E}, \widetilde{H})$. No additional computations are needed.

The contents of this Section is original work of the author, and was originally published in the included paper [PIII], where the error equality is first derived in an abstract setting, and then several different applications are discussed and numerically tested. Functional a posteriori error estimates for mixed approximations of static problems were exposed in [44].

3.5 Computation of the majorants

This Section is dedicated to exposing several different methods to calculate the free functions in the functional majorants. It is clear from the structure of the majorants that the free functions should be chosen as close to the dual variable H as possible. We will first explain in detail the process of global minimization of the majorants. Then we will show computationally cheaper ways to obtain approximations to H, namely, averaging type methods, and post-processing methods.

3.5.1 Global minimization of majorants

Global minimization of a quadratic function $f(Y)$ with respect to Y consists of calculating

$$\{\partial_t f(Y + tq)\}_{t=0} = 0, \tag{3.27}$$

where $t \in \mathbb{R}$, ∂_t is the partial derivative with respect to t, and q is a function form the same space as Y. This technique of obtaining sharp values for functional majorants is widely used, and is exposed in detail in [22] (see also [46]). In the following, we will perform this calculation for all majorants presented in this thesis.

The majorant $\mathcal{M}_E$ and $\mathcal{M}_E^{(1)}$

Globally minimizing the majorant in Theorem 3.3 with respect to the arbitrary function $Y \in H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$ is straightforward since the majorant $\mathcal{M}_E$ is already in a quadratic form. By performing the calculation (3.27) on $\mathcal{M}_E$, and rearranging the terms such that terms related to Y are on the left hand side, one obtains

$$(\mathrm{curl}\, Y, \mathrm{curl}\, q)_{\kappa^{-1}} + (Y, q)_\mu = (F, \mathrm{curl}\, q)_{\kappa^{-1}} - (\widetilde{E}, \mathrm{curl}\, q) + (\mathrm{curl}\, \widetilde{E}, q).$$

By (2.11) the two last terms cancel each other, and we obtain

$$(\mathrm{curl}\, Y, \mathrm{curl}\, q)_{\kappa^{-1}} + (Y, q)_\mu = (F, \mathrm{curl}\, q)_{\kappa^{-1}}. \tag{3.28}$$

Obtaining a value of the majorant of Theorem 3.3 by global minimization consists then of the two steps

1. Calculate an approximation $\widetilde{Y}$ of Y by discretizing (3.28).
2. Calculate the upper bound $\mathcal{M}_E(\widetilde{E}, \widetilde{Y})$.

Note that (3.28) is the weak dual problem (2.33), so by using this procedure $\widetilde{Y}$ is effectively an approximation of the dual variable H, and we could denote $\widetilde{Y}$ by $\widetilde{H}$. Then, by Remark 3.4 we see that we are actually also calculating the exact error of the approximation pair $(\widetilde{E}, \widetilde{H})$ in the combined norm.

The majorant $\mathcal{M}_E^{(\lambda)}$

As a reminder, in the derivation of the majorant of Theorem 3.5 it is assumed that κ is a positive constant. First we transform $\mathcal{M}_E^{(\lambda)}$ into a quadratic form by using the Young inequality (2.6) on the first and third terms of (3.12):

$$\begin{aligned}\mathcal{M}_E^{(\lambda)}(\widetilde{E}, Y) \leq{} & \alpha C_F^2 \|(1-\lambda) r(\widetilde{E}, Y)\|^2 + \frac{1}{\alpha}\|\mathrm{div}\, \widetilde{E}\|^2 + \|\lambda r(\widetilde{E}, Y)\|^2_{\kappa^{-1}} + \\ & + (1+\beta) C_M^2 \overline{\mu} \|(1-\lambda) r(\widetilde{E}, Y)\|^2 + \left(1 + \frac{1}{\beta}\right) \|d(\widetilde{E}, Y)\|^2_\mu \\ ={} & \int_\Omega \left((1-\lambda)^2 C_1 + \lambda^2 \kappa^{-1}\right) r(\widetilde{E}, Y)^2 \, \mathrm{dx} + \\ & + \left(1 + \frac{1}{\beta}\right) \|d(\widetilde{E}, Y)\|^2_\mu + \frac{1}{\alpha}\|\mathrm{div}\, \widetilde{E}\|^2,\end{aligned}$$

where $\alpha, \beta > 0$ arise from the Young inequality, and

$$C_1 := \alpha C_F^2 + (1+\beta) C_M^2 \overline{\mu}. \tag{3.29}$$

It is easy to verify that the optimal form of λ is

$$C_2 := \frac{C_1}{C_1 + \kappa^{-1}} \in (0, 1).$$

By denoting

$$C_3 := \left(1 + \frac{1}{\beta}\right) \tag{3.30}$$

we finally obtain the estimate

$$\begin{aligned} |\!|\!| E - \widetilde{E} |\!|\!|^2 &\leq C_2 \| r(\widetilde{E}, Y) \|^2_{\kappa^{-1}} + C_3 \| d(\widetilde{E}, Y) \|^2_{\mu} + \frac{1}{\alpha} \| \operatorname{div} \widetilde{E} \|^2 \\ &=: \mathcal{M}_E^{(\lambda)}(\widetilde{E}, Y, \alpha, \beta). \end{aligned} \tag{3.31}$$

Globally minimizing the quadratic upper bound (3.31) with respect to $Y \in H(\mathrm{curl})$ results in the following problem:

$$\begin{aligned} &C_2(\operatorname{curl} Y, \operatorname{curl} q)_{\kappa^{-1}} + C_3(Y, q)_{\mu} = \\ &\qquad = C_2(F, \operatorname{curl} q)_{\kappa^{-1}} - C_2(\widetilde{E}, \operatorname{curl} q) + C_3(\operatorname{curl} \widetilde{E}, q), \end{aligned} \tag{3.32}$$

where $q \in H(\mathrm{curl})$. Let $\widetilde{Y}$ be an approximation of Y calculated by discretizing (3.32). Then, for given $\widetilde{E}$ and $\widetilde{Y}$, and for the optimal form of λ, we can easily calculate the optimal values for α and β:

$$\alpha^{opt}(\widetilde{E}, \widetilde{Y}) = \frac{\| \operatorname{div} \widetilde{E} \|}{C_F \| (1 - C_2) r(\widetilde{E}, \widetilde{Y}) \|}, \tag{3.33}$$

$$\beta^{opt}(\widetilde{E}, \widetilde{Y}) = \frac{\| d(\widetilde{E}, \widetilde{Y}) \|_{\mu}}{C_M \overline{\mu}^{1/2} \| (1 - C_2) r(\widetilde{E}, \widetilde{Y}) \|}. \tag{3.34}$$

Note that in (3.33) and (3.34) the constant C_2 depends on C_1 which in turn depends on the *old* values of α and β. Obtaining a value for the majorant of Theorem 3.5 consists then of the following steps:

1. Set $\alpha = \beta = 1$.
2. Calculate an approximation $\widetilde{Y}$ of Y by discretizing (3.32).
3. Update the values of α and β by using (3.33) and (3.34), respectively.
4. Calculate a value for the upper bound $\mathcal{M}_E^{(\lambda)}(\widetilde{E}, \widetilde{Y}, \alpha, \beta)$ by (3.31).

This process is monotone, i.e., repeating the steps 2–4 certain times (or until some stopping criterion is met) produces lower values for the upper bound on every iteration.

The majorant $\mathcal{M}_E^{(0)}$

The global minimization of the majorant $\mathcal{M}_E^{(0)}$ (see Remark 3.1) is very similar to the global minimization of $\mathcal{M}_E^{(\lambda)}$. By transforming $\mathcal{M}_E^{(0)}$ into quadratic form one obtains

$$\begin{aligned} |\!|\!| E - \widetilde{E} |\!|\!|^2 &\leq C_1 \| r(\widetilde{E}, Y) \|^2 + C_3 \| d(\widetilde{E}, Y) \|^2_{\mu} + \frac{1}{\alpha} \| \operatorname{div} \widetilde{E} \|^2 \\ &=: \mathcal{M}_E^{(0)}(\widetilde{E}, Y, \alpha, \beta), \end{aligned} \tag{3.35}$$

where the constants C_1 and C_3 are defined in (3.29) and (3.30). Again, the constants $\alpha, \beta > 0$ arise from the Young inequality. Globally minimizing the quadratic upper bound (3.35) with respect to $Y \in H(\mathrm{curl})$ results in

$$\begin{aligned} C_1(\operatorname{curl} Y, \operatorname{curl} q) + C_3(Y, q)_\mu = \\ = C_1(F, \operatorname{curl} q) - C_1(\widetilde{E}, \operatorname{curl} q)_\kappa + C_3(\operatorname{curl} \widetilde{E}, q), \end{aligned} \tag{3.36}$$

where $q \in H(\mathrm{curl})$. Let $\widetilde{Y}$ be an approximation of Y calculated by discretizing (3.36). Then, for given $\widetilde{E}$ and $\widetilde{Y}$ we can easily calculate the optimal values for α and β:

$$\alpha^{opt}(\widetilde{E}, \widetilde{Y}) = \frac{\|\operatorname{div} \widetilde{E}\|}{C_F \|r(\widetilde{E}, \widetilde{Y})\|}, \tag{3.37}$$

$$\beta^{opt}(\widetilde{E}, \widetilde{Y}) = \frac{\|d(\widetilde{E}, \widetilde{Y})\|_\mu}{C_M \overline{\mu}^{1/2} \|r(\widetilde{E}, \widetilde{Y})\|}. \tag{3.38}$$

So, obtaining a value for the majorant $\mathcal{M}_E^{(0)}$ consists of the following steps:

1. Set $\alpha = \beta = 1$.
2. Calculate an approximation $\widetilde{Y}$ of Y by discretizing (3.36).
3. Update the values of α and β by using (3.37) and (3.38), respectively.
4. Calculate a value for the upper bound $\mathcal{M}_E^{(0)}(\widetilde{E}, \widetilde{Y}, \alpha, \beta)$ by (3.35).

As before, this process is monotone, i.e., repeating the steps 2–4 certain times (or until some stopping criterion is met) produces lower values for the upper bound on every iteration.

3.5.2 Averaging type methods

The so-called gradient averaging type methods originate from [47] and [48]. These type methods are also called gradient recovery methods, and error indicators arising from these methods are commonly called Zienkiewicz-Zhu (ZZ) error estimators. For the eddy current problem, a recovery procedure and a ZZ type error estimator was presented in [32].

The advantage of averaging procedures is that they provide a computationally cheap way to obtain approximations to the dual variable H. The disadvantage is the fact that they are relatively inaccurate. We define here two different averaging operators. For this we will assume that the approximation to the primal problem (2.32) is calculated by linear Nédélec elements of the first family introduced in [29]. We will denote this finite element solution by E_h. Then, the curl of E_h is constant in each element. The following averaging procedures are done with respect to the patches ω_{N} and ω_{E} (see Figure 2).

The nodal averaging operator G_{N} is defined by

$$G_{\mathrm{N}} E_h(\mathrm{N}) = \begin{cases} 0 & \text{if } \mathrm{N} \in \Gamma_N \\ \sum_{\mathrm{T} \in \omega_{\mathrm{N}}} \frac{|\mathrm{T}|}{|\omega_{\mathrm{N}}|} \left(\{\!\{\mu\}\!\}_{\mathrm{T}}^{-1} \operatorname{curl} E_h \right) \Big|_{\mathrm{T}} & \text{otherwise} \end{cases}, \tag{3.39}$$

which defines the value of $G_{\mathrm{N}} E_h$ at the node N. Then, the averaged function $G_{\mathrm{N}} E_h$ is defined by piecewise affine extension on the whole domain. Here $\{\!\{\mu\}\!\}_{\mathrm{T}}$ is the average of μ on the element T. The recovery procedure proposed in [32] is very similar to (3.39).

Nodal averaging is a natural choice if $d = 2$, since the dual variable belongs to $H^1_{0,\Gamma_N}(\Omega)$. However, if $d = 3$, the dual variable belongs to $H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$, but using (3.39) produces functions from $H^1_{0,\Gamma_N}(\Omega, \mathbb{R}^3)$. Therefore, we propose the following edge averaging operator G_E by

$$G_\mathrm{E} E_h(\mathrm{E}) = \begin{cases} 0 & \text{if } \mathrm{N} \in \Gamma_N \\ \frac{|\mathrm{E}|}{\#\omega_\mathrm{E}}\, t_\mathrm{E} \cdot \sum_{\mathrm{T}\in\omega_\mathrm{E}} \left(\{\!\{\mu\}\!\}_\mathrm{E}^{-1} \mathrm{curl}\, E_h \right)\Big|_\mathrm{T} & \text{otherwise} \end{cases}, \tag{3.40}$$

which defines the value of $G_\mathrm{E} E_h$ at the edge E. Then, the averaged function $G_\mathrm{E} u_h$ is defined by an extension with the help of the linear Nédélec elements. Here $\{\!\{\mu\}\!\}_\mathrm{E}$ is the average of μ on E, and t_E is the tangential unit vector to E.

3.5.3 Post-processing

If $d = 2$ a function defined by (3.39) can be represented by

$$Y_h = \sum_{\mathrm{N}\in\mathcal{N}_h} c_\mathrm{N} \phi_\mathrm{N}, \tag{3.41}$$

where $c_\mathrm{N} = G_\mathrm{N} E_h(\mathrm{N})$ and ϕ_N is the global linear Courant basis function related to the node N (see, e.g., [8]). Since ϕ_N is nonzero only in the patch ω_N, we can easily derive a local optimization routine to obtain more accurate approximations of the dual variable H. This involves minimizing the majorant (3.1) in all patches ω_N: for all N we update the degree of freedom c_N by minimizing

$$\begin{aligned} M_\mathrm{N} &:= \mathcal{M}_E(E_h, Y_h)\big|_{\omega_\mathrm{N}} \\ &= \Big\| F - \kappa E_h - \sum_{\mathrm{N}'\in\mathcal{N}_h(\omega_\mathrm{N})} c_{\mathrm{N}'} \underline{\mathrm{curl}}\, \phi_{\mathrm{N}'} \Big\|^2_{\omega_\mathrm{N},\kappa^{-1}} + \Big\| \sum_{\mathrm{N}'\in\mathcal{N}_h(\omega_\mathrm{N})} c_{\mathrm{N}'} \phi_{\mathrm{N}'} - \mu^{-1} \mathrm{curl}\, E_h \Big\|^2_{\omega_\mathrm{N},\mu}, \end{aligned}$$

where $\mathcal{N}_h(\omega_\mathrm{N})$ denotes all the nodes in the element contained in the patch ω_N. By calculating $\partial_{c_\mathrm{N}} M_\mathrm{N} = 0$ we obtain a new value c^*_N:

$$P_\mathrm{N} Y_h(\mathrm{N}) = c^*_\mathrm{N} = \begin{cases} 0 & \text{if } \mathrm{N} \in \Gamma_N \\ A_\mathrm{N}/B_\mathrm{N} & \text{otherwise} \end{cases}, \tag{3.42}$$

where

$$\begin{aligned} A_\mathrm{N} = -\Big(F - \kappa E_h - \sum_{\mathrm{N}'\in\mathcal{N}_h(\omega_\mathrm{N})\backslash\mathrm{N}} c_{\mathrm{N}'} \underline{\mathrm{curl}}\, \phi_{\mathrm{N}'}, \underline{\mathrm{curl}}\, \phi_\mathrm{N} \Big)_{\omega_\mathrm{N},\kappa^{-1}} + \\ + \Big(\sum_{\mathrm{N}'\in\mathcal{N}_h(\omega_\mathrm{N})\backslash\mathrm{N}} c_{\mathrm{N}'} \phi_{\mathrm{N}'} - \mu^{-1} \mathrm{curl}\, E_h, \phi_\mathrm{N} \Big)_{\omega_\mathrm{N},\mu} \end{aligned}$$

and

$$B_\mathrm{N} = -(\underline{\mathrm{curl}}\, \phi_\mathrm{N}, \underline{\mathrm{curl}}\, \phi_\mathrm{N})_{\omega_\mathrm{N},\kappa^{-1}} - (\phi_\mathrm{N}, \phi_\mathrm{N})_{\omega_\mathrm{N},\mu}.$$

The function whose degrees of freedom are updated for all nodes by (3.42) is again in the form of (3.41), so the procedure P_{N} can be applied to itself multiple times.

Similarly, if $d = 3$, a function defined by (3.40) can be represented by the sum $Y_h = \sum_{\mathrm{E} \in \mathcal{E}_h} c_{\mathrm{E}} \phi_{\mathrm{E}}$, where $c_{\mathrm{E}} = G_{\mathrm{E}} E_h(\mathrm{E})$ and ϕ_{E} is the global linear Nédélec basis function related to the edge E (see, e.g., [26,29]). By minimizing $\mathcal{M}_E(E_h, Y_h)\big|_{\omega_{\mathrm{E}}}$ we obtain a new value c^*_{E}:

$$P_{\mathrm{E}} Y_h(\mathrm{E}) = c^*_{\mathrm{E}} = \begin{cases} 0 & \text{if } \mathrm{E} \in \Gamma_N \\ A_{\mathrm{E}}/B_{\mathrm{E}} & \text{otherwise} \end{cases}, \tag{3.43}$$

where

$$A_{\mathrm{E}} = -\Big(F - \kappa E_h - \sum_{\mathrm{E}' \in \mathcal{E}_h(\omega_{\mathrm{E}}) \setminus \mathrm{E}} c_{\mathrm{E}'} \operatorname{curl} \phi_{\mathrm{E}'}, \operatorname{curl} \phi_{\mathrm{E}}\Big)_{\omega_{\mathrm{E}}, \kappa^{-1}} + \\ + \Big(\sum_{\mathrm{E}' \in \mathcal{E}_h(\omega_{\mathrm{E}}) \setminus \mathrm{E}} c_{\mathrm{E}'} \phi_{\mathrm{E}'} - \mu^{-1} \operatorname{curl} E_h, \phi_{\mathrm{E}}\Big)_{\omega_{\mathrm{E}}, \mu}$$

and

$$B_{\mathrm{E}} = -(\operatorname{curl} \phi_{\mathrm{E}}, \operatorname{curl} \phi_{\mathrm{E}})_{\omega_{\mathrm{E}}, \kappa^{-1}} - (\phi_{\mathrm{E}}, \phi_{\mathrm{E}})_{\omega_{\mathrm{E}}, \mu}.$$

Similarly as for the post-processing operator P_{N}, the operator P_{E} can be applied to itself multiple times.

It should be noted that parallelization of the operators P_{N} and P_{E} has to be done with care. Parallel calculations of (3.42) and (3.43) have to be done in patches which do not overlap each other.

This type post-processing techniques for fluxes were exposed in the included article [PIV] (and also in the conference article [2]).

3.6 Numerical examples

In this section we test, with some simple academic test examples, all the estimates and the equality derived in this chapter. We will use the following data:

Data 1. We choose the unit square $\Omega := (0,1)^2$ with $\kappa = 10^{-1}$ and $\mu = 1$. The continuous exact solution and the right hand side are

$$E := \begin{bmatrix} \sin(\pi x_2) \\ \sin(\pi x_1) \end{bmatrix}, \quad F := (\pi^2 + \kappa) \begin{bmatrix} \sin(\pi x_2) \\ \sin(\pi x_1) \end{bmatrix}.$$

Obviously E satisfies zero Dirichlet boundary conditions on the whole boundary, i.e., $\Gamma_N = \varnothing$ and $\Gamma_D = \partial\Omega$. It is clear that E belongs to $H_0(\operatorname{curl}, \Omega)$, and it even belongs to $H^1_0(\Omega, \mathbb{R}^2)$. Moreover, $H := \operatorname{curl} E = \pi(\cos(\pi x_1) - \cos(\pi x_2))$ belongs to $H^1(\Omega)$. The exact solution is visualized in Figure 3.

Data 2. We choose the unit square $\Omega := (0,1)^2$ with $\kappa = \mu = 1$. We split the domain in two parts $\Omega_1 := \{x \in \Omega \mid x_1 > x_2\}$ and $\Omega_2 := \Omega \backslash \overline{\Omega}_1$ in order to define the following discontinuous exact solution:

$$E\big|_{\Omega_1}(x) := \begin{bmatrix} \sin(2\pi x_1) + 2\pi\cos(2\pi x_1)(x_1 - x_2) \\ \sin\left((x_1 - x_2)^2(x_1 - 1)^2 x_2\right) - \sin(2\pi x_1) \end{bmatrix}, \quad E\big|_{\Omega_2}(x) := 0.$$

Since on $\Gamma_/ := \{x \in \Omega \mid x_1 = x_2\}$ we have

$$\begin{aligned} E\big|_{\Omega_1} \times n &= \frac{1}{2^{1/2}}\left(2\pi\cos(2\pi x_1)(x_1 - x_2) + \sin\left((x_1 - x_2)^2(x_1 - 1)^2 x_2\right)\right), \\ E\big|_{\Omega_2} \times n &= 0, \end{aligned}$$

we see that $E\big|_{\Omega_1} \times n = 0$ on $\Gamma_/$. We conclude that the tangential component is continuous on $\Gamma_/$, so E belongs to $H(\operatorname{curl}, \Omega)$. Moreover,

$$\begin{aligned} \operatorname{curl} E\big|_{\Omega_1} &= 2x_2(x_1 - x_2)(x_1 - 1)(2x_1 - x_2 - 1)\cos\left(x_2(x_1 - x_2)^2(x_1 - 1)^2\right), \\ \operatorname{curl} E\big|_{\Omega_2} &= 0, \end{aligned}$$

and clearly $\operatorname{curl} E\big|_{\Omega_1} = 0$ on $\Gamma_/$, i.e., $\operatorname{curl} E$ is continuous on $\Gamma_/$. Also, it is easy to see that $\operatorname{curl} E$ vanishes on the whole boundary, so $H := \operatorname{curl} E \in H^1_0(\Omega)$. Thus, the exact solution satisfies zero Neumann boundary conditions on the whole boundary, i.e., $\Gamma_D = \varnothing$ and $\Gamma_N = \partial\Omega$. The exact solution is visualized in Figure 4.

Data 3. We choose the L-shaped domain $\Omega := (0,1)^2 \backslash \left([1/2, 1] \times [0, 1/2]\right)$ with

$$\kappa = 1, \quad \mu = 10^3 \quad \text{and} \quad F = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

We set zero Dirichlet boundary conditions on the whole boundary, i.e., $\Gamma_N = \varnothing$ and $\Gamma_D = \partial\Omega$. The exact solution of this problem is unknown.

Data 4. We choose the unit cube $\Omega := (0,1)^3$ with $\kappa = \mu = 1$. We split the domain in two parts $\Omega_1 := \{x \in \Omega \mid x_1 > x_2\}$ and $\Omega_2 = \Omega \backslash \overline{\Omega}_1$ in order to define the following discontinuous exact solution:

$$E(x) := \chi_{\Omega_1}(x) \begin{bmatrix} \sin(2\pi x_1) + 2\pi\cos(2\pi x_1)(x_1 - x_2) \\ \sin\left((x_1 - x_2)^2(x_1 - 1)^2 x_2\right) - \sin(2\pi x_1) \\ 0 \end{bmatrix} + E^{(1)} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

where χ denotes the characteristic function, and $E^{(1)} := \prod_{i=1}^3 x_i^2(1 - x_i)^2$. Thus, we extended the discontinuous vector field of Data 2 by zero in the third component, and added a smooth bubble in the third component. Since on the plane $\Gamma_/ := \{x \in \Omega \mid x_1 = x_2\}$ we have

$$\begin{aligned} E\big|_{\Omega_1} \times n &= \frac{1}{2^{1/2}} \begin{bmatrix} E^{(1)} \\ E^{(1)} \\ 2\pi\cos(2\pi x_1)(x_1 - x_2) + \sin\left((x_1 - x_2)^2(x_1 - 1)^2 x_2\right) \end{bmatrix}, \\ E\big|_{\Omega_2} \times n &= \frac{1}{2^{1/2}} \begin{bmatrix} E^{(1)} \\ E^{(1)} \\ 0 \end{bmatrix}, \end{aligned}$$

we see that $E\big|_{\Omega_1} \times n = \frac{1}{2^{1/2}}(E^{(1)}, E^{(1)}, 0)^T$ on $\Gamma_/$. We conclude that the tangential component is continuous on $\Gamma_/$, so E belongs to $H(\mathrm{curl}, \Omega)$. Moreover,

$$\mathrm{curl}\, E = \chi_{\Omega_1}(x) \begin{bmatrix} 0 \\ 0 \\ E^{(2)}(x) \end{bmatrix} + \begin{bmatrix} \partial_2 E^{(1)}(x) \\ -\partial_1 E^{(1)}(x) \\ 0 \end{bmatrix},$$

where

$$E^{(2)}(x) := 2x_2(x_1 - x_2)(x_1 - 1)(2x_1 - x_2 - 1)\cos\big(x_2(x_1 - x_2)^2(x_1 - 1)^2\big).$$

On $\Gamma_/$ we have

$$\mathrm{curl}\, E\big|_{\Omega_1} \times n = \frac{1}{2^{1/2}} \begin{bmatrix} E^{(2)}(x) \\ E^{(2)}(x) \\ \partial_1 E^{(1)}(x) - \partial_2 E^{(1)}(x) \end{bmatrix},$$

$$\mathrm{curl}\, E\big|_{\Omega_2} \times n = \frac{1}{2^{1/2}} \begin{bmatrix} 0 \\ 0 \\ \partial_1 E^{(1)}(x) - \partial_2 E^{(1)}(x) \end{bmatrix}.$$

Since $E^{(2)}(x) = 0$ on $\Gamma_/$, we see that the tangential component of curl E is continuous on $\Gamma_/$. It is also easy to verify that curl $E \times n = 0$ on the whole boundary, so $H := \mathrm{curl}\, E \in H_0(\mathrm{curl}, \Omega)$. Thus, the exact solution satisfies zero Neumann boundary conditions on the whole boundary, i.e., $\Gamma_D = \varnothing$ and $\Gamma_N = \partial\Omega$.

Data 5. We choose the L-shaped domain $\Omega := (0,1)^3 \backslash [1/2, 1]^3$. We define for $i = 1, 2, 3$ the subdomains $\Omega_i := \{x \in \Omega \mid x_i > 1/2\}$ in order to define the following discontinuous data:

$$\kappa\big|_{\Omega_1} = \mu\big|_{\Omega_2} = 1, \quad \kappa\big|_{\Omega\backslash\overline{\Omega}_1} = \mu\big|_{\Omega\backslash\overline{\Omega}_2} = 10^2,$$
$$F\big|_{\Omega_3} = [1, 0, 0]^T, \quad F\big|_{\Omega\backslash\overline{\Omega}_3} = [0, 0, 1]^T.$$

We set zero Neumann boundary conditions on the whole boundary, i.e., $\Gamma_D = \varnothing$ and $\Gamma_N = \partial\Omega$. The exact solution of this problem is unknown.

The right hand sides generated by the exact solutions of Data 2 and 4 are not printed here since they are simply too long. In fact, the right hand sides for these problems were automatically calculated from the exact solutions by code created by the author. This code utilizes the `Symbolic Math Toolbox` for `MATLAB`.

In all the numerical examples we solve the primal problem (2.32) with linear Nédélec elements of the first family (see, e.g., [26, 29]) for $d \geq 2$. If $d = 3$ we will solve also the dual problem (2.33), and the problems resulting from global minimization of the majorant (see Section 3.5), with linear Nédélec elements of the first family. If $d = 2$, we will solve those problems with the linear Courant elements (see, e.g., [8]). A finite element discretization will be denoted by lowercase h, i.e., (E_h, H_h). If not otherwise stated, the resulting linear systems are solved by a direct method.

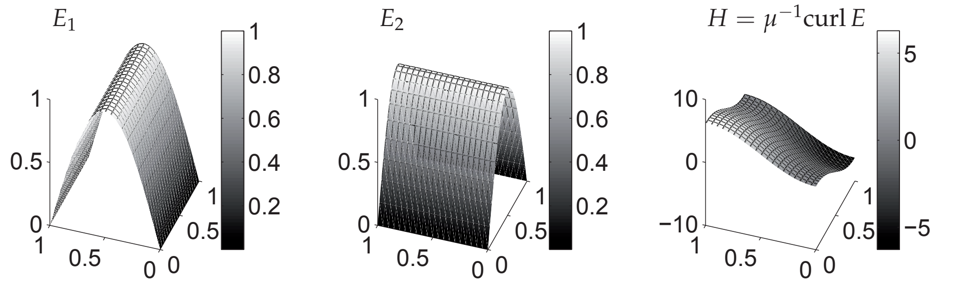


FIGURE 3 The exact solution (E, H) of Data 1.

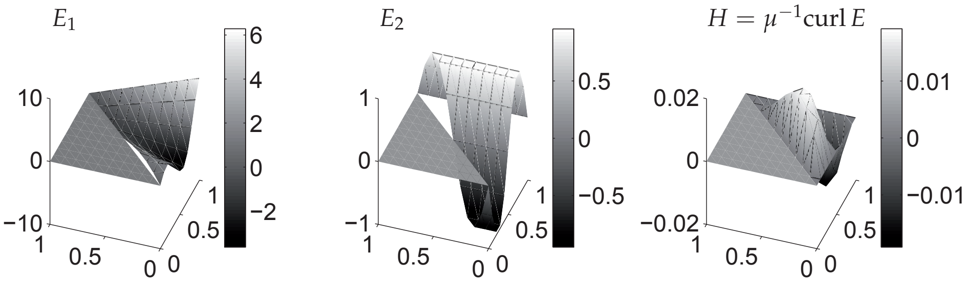


FIGURE 4 The exact solution (E, H) of Data 2.

We will not test the majorants of Theorems 3.3 and 3.4 separately, since we would in any case choose $Y = H_h$ and $X = E_h$, and by Remark 3.4 the majorants then give the exact error in the combined norm.

In the case of the advanced form of the majorant of Theorem 3.5 we will post-process the finite element approximation $E_h \in H_0(\text{curl}, \Omega)$ into $H_0^1(\Omega)$ by a simple node averaging procedure which we will denote by Q.

In all the tables presented here, the values are cut-offs from the values reported by MATLAB. No rounding is done.

We will include in some of the numerical results also values of the so-called efficiency indexes. For the minorants we denote

$$I_{E,\ominus} = \frac{m_E(E_h, Z)^{1/2}}{|\!|\!| E - E_h |\!|\!|} \leq 1, \qquad I_{H,\ominus} = \frac{m_H(H_h, \hat{Z})^{1/2}}{[\![H - H_h]\!]} \leq 1,$$

and for the advanced form of the majorant we denote

$$I_{E,\oplus}^{(\lambda)} = \frac{\left(\mathcal{M}_E^{(\lambda)}(Q(E_h), Y)\right)^{1/2}}{|\!|\!| E - Q(E_h) |\!|\!|} \geq 1.$$

The closer these values are to 1, the better estimates we have. For the error equality of Theorem 3.6 the efficiency index is always 1 (aside from possible errors arising from using integration quadratures, or the machine precision of floating point arithmetic).

Example 1. We will test the minorants. In order to obtain a value for $m_E(\widetilde{E}, Z)$ we need to essentially choose Z such that it is a better approximation to the primal variable E than $\widetilde{E}$. The same holds for the minorant m_H. In practice we will compute approximations in subsequently refined meshes $\mathcal{T}_{h_1}, \mathcal{T}_{h_2}, \ldots$ and obtain the corresponding approximations $E_{h_1}, E_{h_2}, \ldots$ and $H_{h_1}, H_{h_2}, \ldots$. The values of the minorants for E_{h_i} and H_{h_i} are then obtained with a "one step delay" by $m_E(E_{h_i}, E_{h_{i+1}})$ and $m_H(H_{h_i}, H_{h_{i+1}})$, respectively. In this example in each refinement step we refine the whole mesh. Tables 1 and 2 contain the results for Data 1 and 4 respectively. The efficiency indexes vary between 0.76 and 0.86, which indicates good performance.

TABLE 1 Example 1 ($d = 2$, Data 1) The values of the minorants.

$\#\mathcal{T}_{h_i}$	$\vert\vert\vert E - E_{h_i} \vert\vert\vert$	$m_E(E_{h_i}, E_{h_{i+1}})^{1/2}$	$I_{E,\ominus}$	$\vert\vert[H - H_{h_i}]\vert\vert$	$m_H(H_{h_i}, H_{h_{i+1}})^{1/2}$	$I_{H,\ominus}$
200	0.18110	0.15671	0.86	2.80850	2.42780	0.86
800	0.09076	0.07859	0.86	1.41190	1.22210	0.86
3200	0.04541	0.03932	0.86	0.70714	0.61230	0.86
12800	0.02278	0.01966	0.86	0.35374	0.30633	0.86
51200	0.01135	0.00983	0.86	0.17690	0.15319	0.86
204800	0.00567	-	-	0.08845	-	-

TABLE 2 Example 1 ($d = 3$, Data 4) The values of the minorants.

$\#\mathcal{T}_{h_i}$	$\vert\vert\vert E - E_{h_i} \vert\vert\vert$	$m_E(E_{h_i}, E_{h_{i+1}})^{1/2}$	$I_{E,\ominus}$	$\vert\vert[H - H_{h_i}]\vert\vert$	$m_H(H_{h_i}, H_{h_{i+1}})^{1/2}$	$I_{H,\ominus}$
384	0.72000	0.55085	0.76	0.063734	0.052376	0.82
3072	0.46365	0.37490	0.80	0.036315	0.031014	0.85
24576	0.27280	0.22879	0.83	0.018891	0.016311	0.86
196608	0.14856	-	-	0.009528	-	-

Example 2. We test the advanced form of the majorant $\mathcal{M}_E^{(\lambda)}(Q(E_h), Y)$ with various values of κ for Data 1 on a mesh with 3200 elements. We compare the performance also with the special cases of Remark 3.1, i.e., we choose $\lambda = 1$ and $\lambda = 0$. In Table 3 are results where the free functions Y were chosen by globally minimizing the corresponding majorants. In Table 4 are results where the free functions Y were all chosen as the averaged function $G_{\mathrm{N}}E_h$ (see Subsection 3.5.2). The special case $\lambda = 1$ is sensitive for small values of κ, and the special case $\lambda = 0$ is sensitive for big values of κ. It can be concluded that the advanced form of the majorant gives a good error bound with all values of κ.[2]

TABLE 3 Example 2 ($d = 2$, Data 1) The efficiency index values of the majorants with various values of κ on a mesh with 3200 elements. The Y were chosen by global minimization.

κ	$I_{E,\oplus}^{(1)}$	$I_{E,\oplus}^{(0)}$	$I_{E,\oplus}^{(\lambda)}$
10^{-4}	248.61	2.63	2.63
10^{-3}	78.62	2.63	2.63
10^{-2}	24.88	2.63	2.62
10^{-1}	7.92	2.63	2.59
10^{0}	2.67	2.62	2.30
10^{1}	1.26	2.62	1.26
10^{2}	1.02	3.54	1.02
10^{3}	1.00	10.1	1.00
10^{4}	1.00	31.81	1.00

TABLE 4 Example 2 ($d = 2$, Data 1) The efficiency index values of the majorants with various values of κ on a mesh with 3200 elements. All Y were chosen as the averaged function $G_{\mathrm{N}}E_h$.

κ	$I_{E,\oplus}^{(1)}$	$I_{E,\oplus}^{(0)}$	$I_{E,\oplus}^{(\lambda)}$
10^{-4}	1778.86	8.39	8.39
10^{-3}	562.52	8.39	8.39
10^{-2}	177.88	8.39	8.39
10^{-1}	56.25	8.39	8.34
10^{0}	17.79	8.38	7.93
10^{1}	5.65	8.31	5.41
10^{2}	1.89	7.82	1.89
10^{3}	1.05	10.81	1.05
10^{4}	1.00	31.92	1.00

[2] This test is also done in [PI] with a smaller mesh. We note that in [PI, equation (2.2)] we incorrectly use the Friedrichs constant C_F. This constant should be the Poincaré constant C_P. In the numerical results of this paper the bound (2.16) is then incorrectly used for both of the constants, so the values in [PI, Tables 1 and 3] are slightly optimistic.

Example 3. We will test the averaging and post-processing operators of Subsections 3.5.2 and 3.5.3 for calculating approximations of the dual variable. The approximation E_h was solved with linear Nédélec elements. For Data 1 and 2 ($d = 2$) we will test the nodal operators G_{N} and P_{N}, and for Data 4 ($d = 3$) we will test the edge operators G_{E} and P_{E}. For comparison we also calculated an finite element approximation H_h to the dual variable. Results for Data 1, 2, and 4 are in Tables 5, 6, and 7, respectively. In all cases only a few post-processing iterations are needed to get relatively close to H_h.

TABLE 5 Example 3 ($d = 2$, Data 1) Errors $|\!|\!|[\, H - \widetilde{H} \,]|\!|\!| \;/\; |\!|\!|[\, H \,]|\!|\!|$ for various approximations $\widetilde{H}$ on three meshes.

$\#\mathcal{T}_h$	200	3200	51200
$G_{\mathrm{N}}E_h$	0.1148	0.0472	0.0220
$P_{\mathrm{N}}^1 G_{\mathrm{N}}E_h$	0.0825	0.0241	0.0086
$P_{\mathrm{N}}^2 G_{\mathrm{N}}E_h$	0.0795	0.0213	0.0064
$P_{\mathrm{N}}^3 G_{\mathrm{N}}E_h$	0.0787	0.0205	0.0058
$P_{\mathrm{N}}^4 G_{\mathrm{N}}E_h$	0.0784	0.0202	0.0054
$P_{\mathrm{N}}^5 G_{\mathrm{N}}E_h$	0.0782	0.0200	0.0053
H_h	0.0779	0.0196	0.0049

TABLE 6 Example 3 ($d = 2$, Data 2) Errors $|\!|\!|[\, H - \widetilde{H} \,]|\!|\!| \;/\; |\!|\!|[\, H \,]|\!|\!|$ for various approximations $\widetilde{H}$ on three meshes.

$\#\mathcal{T}_h$	200	3200	51200
$G_{\mathrm{N}}E_h$	0.4743	0.1679	0.0730
$P_{\mathrm{N}}^1 G_{\mathrm{N}}E_h$	0.3547	0.0968	0.0291
$P_{\mathrm{N}}^2 G_{\mathrm{N}}E_h$	0.3411	0.0885	0.0229
$P_{\mathrm{N}}^3 G_{\mathrm{N}}E_h$	0.3379	0.0868	0.0219
$P_{\mathrm{N}}^4 G_{\mathrm{N}}E_h$	0.3363	0.0862	0.0217
$P_{\mathrm{N}}^5 G_{\mathrm{N}}E_h$	0.3354	0.0858	0.0215
H_h	0.3328	0.0845	0.0211

TABLE 7 Example 3 ($d = 3$, Data 4) Errors $|\!|\!|[\, H - \widetilde{H} \,]|\!|\!| \;/\; |\!|\!|[\, H \,]|\!|\!|$ for various approximations $\widetilde{H}$ on three meshes.

$\#\mathcal{T}_h$	384	3072	24576
$G_{\mathrm{E}}E_h$	1.5495	0.8770	0.4668
$P_{\mathrm{E}}^1 G_{\mathrm{E}}E_h$	0.9977	0.5644	0.2880
$P_{\mathrm{E}}^2 G_{\mathrm{E}}E_h$	0.8784	0.5001	0.2518
$P_{\mathrm{E}}^3 G_{\mathrm{E}}E_h$	0.8231	0.4681	0.2369
$P_{\mathrm{E}}^4 G_{\mathrm{E}}E_h$	0.7911	0.4484	0.2285
$P_{\mathrm{E}}^5 G_{\mathrm{E}}E_h$	0.7712	0.4351	0.2229
H_h	0.7127	0.3833	0.1945

Example 4. We will numerically verify the equality of Theorem 3.6 for Data 4. In this test the main quantity of interest is the difference ε between the exact error and the value given by the majorant for some approximation $(\widetilde{E}, \widetilde{H})$, i.e., $\varepsilon := \left| |[(E,H) - (\widetilde{E}, \widetilde{H})]| - \mathcal{M}(\widetilde{E}, \widetilde{H})^{1/2} \right|$. In Table 8 we have calculated approximations (E_h, H_h) by solving the linear systems directly. In Table 9 the linear systems were solved by an iterative method where the stopping tolerance was set to the crude value of 10^{-6}. These approximations are denoted by (E_{iter}, H_{iter}). With this setting the iterative method did not converge, and the errors of the approximations actually grow as the mesh size was increased. This was done in purpose to obtain approximations which are relatively far from having the Galerkin orthogonality property. In Table 10 the approximation E_h for the primal variable was calculated by solving the linear system directly. The approximation for the dual variable was calculated by $G_{\mathrm{E}} E_h$, i.e., by averaging curl E_h to edges (see Subsection 3.5.2). In the numerical results performed non-zero values of the difference ε was between 10^{-16}–10^{-15}, which is in the limit of machine precision. Thus these values are considered zero.

TABLE 8 Example 4 ($d = 3$, Data 4) Linear systems solved directly.

$\#\mathcal{T}_h$	$\|[(E,H) - (E_h, H_h)]\|$	$\mathcal{M}(E_h, H_h)^{1/2}$	difference ε
384	0.7228185218	0.7228185218	3.330669074e-16
3072	0.3717887807	0.3717887807	6.106226635e-16
24576	0.1883612515	0.1883612515	2.775557562e-16
196608	0.0945757836	0.0945757836	8.604228441e-16

TABLE 9 Example 4 ($d = 3$, Data 4) Linear systems solved with an iterative method.

$\#\mathcal{T}_h$	$\|[(E,H) - (E_{iter}, H_{iter})]\|$	$\mathcal{M}(E_{iter}, H_{iter})^{1/2}$	difference ε
384	0.909324998	0.909324998	1.110223025e-16
3072	1.368867273	1.368867273	2.220446049e-16
24576	2.319303789	2.319303789	8.437694987e-15
196608	3.586832702	3.586832702	8.881784197e-16

TABLE 10 Example 4 ($d = 3$, Data 4) Approximation for the dual variable obtained by averaging.

$\#\mathcal{T}_h$	$\|[(E,H) - (E_h, G_{\mathrm{E}} E_h)]\|$	$\mathcal{M}(E_h, G_{\mathrm{E}} E_h)^{1/2}$	difference ε
384	0.7332143085	0.7332143085	2.220446049e-16
3072	0.3784213896	0.3784213896	1.165734176e-15
24576	0.1921457906	0.1921457906	1.276756478e-15
196608	0.0972059791	0.0972059791	6.661338148e-16

The equality was verified with various other test data in both dimensions $d = 2$ and $d = 3$. In all cases the values of ε were within the limit of machine precision. In [PIII] we numerically verify the equality also in the context of the reaction diffusion problem.

Next we will concentrate on investigating the error indication properties of the majorant. We have already established that the majorant gives the *global* error value of a given approximation $(\widetilde{E}, \widetilde{H})$. In order to perform adaptive refinement of element meshes, one needs to be able to approximate the error distribution in the domain. In the next examples we compare optimal refinement achieved by using the exact error distribution e_{T} to the refinement provided by the distribution of the majorant η_{T}, where

$$\begin{aligned} e_{\mathrm{T}}^2 &:= \|[(E,H) - (\widetilde{E},\widetilde{H})]\|_{\mathrm{T}}^2 := \|\!| E - \widetilde{E} \|\!|_{\mathrm{T}}^2 + [\![H - \widetilde{H}]\!]_{\mathrm{T}}^2, \\ \eta_{\mathrm{T}}^2 &:= \mathcal{M}(\widetilde{E},\widetilde{H})\Big|_{\mathrm{T}} := \|F - \kappa\widetilde{E} - \operatorname{curl}\widetilde{H}\|_{\mathrm{T},\kappa^{-1}}^2 + \|\widetilde{H} - \mu^{-1}\operatorname{curl}\widetilde{E}\|_{\mathrm{T},\mu}^2, \end{aligned} \tag{3.44}$$

and $\mathrm{T} \in \mathcal{T}_h$ denotes an element of the mesh discretization.

In order to measure the accuracy of the error indication process, we define the following *strong* measure (see [22, Section 2.1.2]): the indicator η_{T} is said to be ε-accurate if

$$\Theta_{strong}(\eta_{\mathrm{T}}) := \frac{\|e_{\mathrm{T}} - \eta_{\mathrm{T}}\|_{\mathbb{R}^d}}{\|e_{\mathrm{T}}\|_{\mathbb{R}^d}} \leq \varepsilon.$$

If ε is small, we can say (in some strong sense) that, the error indicator η_{T} is accurate.

In the following examples we will in each step refine 30% of elements with the highest amount of error. The refinement of element meshes is done by regular refinement such that the resulting mesh does not contain hanging nodes. This marking of elements will be denoted by $\mathbb{k}$. The marker $\mathbb{k}$ produces a vector containing ones for elements to be refined, and zeros for elements not to be refined. We define the following *weak* measure (see [22, Section 2.1.2]): the indicator η_{T} is said to be ε-accurate with respect to the marker $\mathbb{k}$ if

$$\Theta_{weak}(\eta_{\mathrm{T}}, \mathbb{k}) := 1 - \frac{\sum(\mathbb{k}(e_{\mathrm{T}}) \wedge \mathbb{k}(\eta_{\mathrm{T}}))}{0.3 \cdot \#\mathcal{T}_h} \leq \varepsilon,$$

where $\wedge$ is the logical `AND` operator. Again, if ε is small, we can say (in some weak sense) that, the error indicator η_{T} is accurate. The measure $\Theta_{weak}(\eta_{\mathrm{T}}, \mathbb{k})$ has been normalized in such a way that it obtains a value 0 when the markers are identical and 1 if they are completely different.

It is clear that a considerably larger class of error indicators may be accurate in the weak sense, while not being accurate in the strong sense.

Example 5. We solve the primal and dual problem pair (E_h, H_h) in adaptively refined meshes for Data 1. We compare optimal refinement achieved by using the exact error distribution e_{T} to the refinement provided by the distribution of the majorant η_{T}. We start from a regular mesh with 200 elements, and perform nine refinement iterations, where on each iteration 30% of elements with the highest amount of error are refined. The converge histories are depicted in Figure 5. In Figure 6 we have depicted the meshes after the fourth refinement. Figure 7 depicts one of the finest parts of the final meshes. We see from Table 11 that the number of elements in the optimal meshes and the meshes produced using η_T are very close to each other. In fact, adaptive refinement using η_T is very close to optimal in each step, and the resulting approximation after the last refinement is practically the same. In the last mesh the combined error of the approximation pair was 0.1253981450 in the optimal mesh, and 0.1255987478 in the mesh generated by η_{T}. Table 12 shows the error indicator measures with respect to the optimal meshes. Both the strong and weak measures of the accuracy of the error indicator are very good.

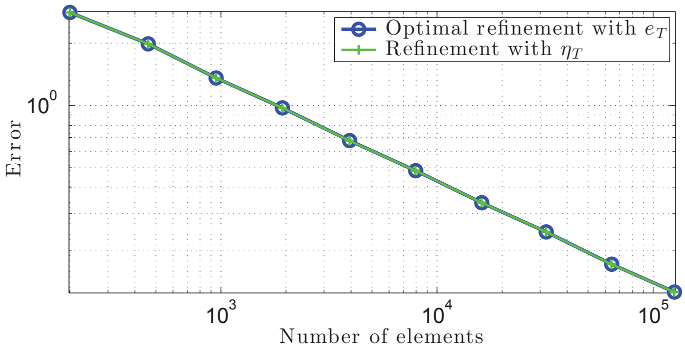


FIGURE 5 Example 5 ($d = 2$, Data 1) Adaptive computation where the error is measured in the combined norm.

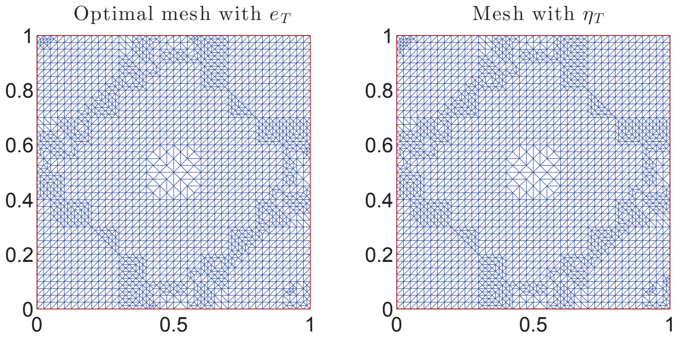


FIGURE 6 Example 5 ($d = 2$, Data 1) Adaptive mesh after the fourth refinement.

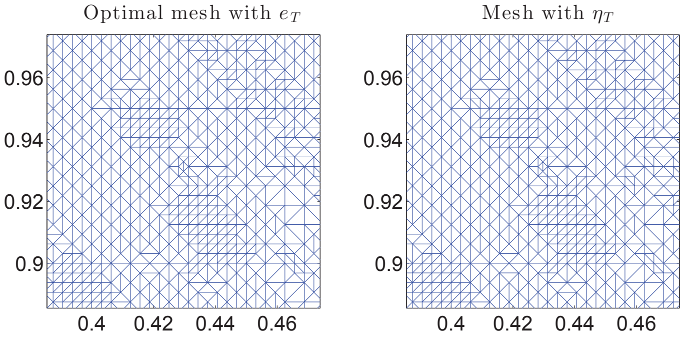


FIGURE 7 Example 5 ($d = 2$, Data 1) One of the most fine parts in the final adaptive mesh.

TABLE 11 Example 5 ($d = 2$, Data 1) The number of elements in the optimal meshes and the meshes generated by the help of η_T.

Ref.	$\#\mathcal{T}_h$ optimal	$\#\mathcal{T}_h$ with η_T	difference	difference %
0	200	200	0	0
1	461	461	0	0
2	950	950	0	0
3	1926	1926	0	0
4	3937	3935	2	0.05
5	7963	7940	23	0.28
6	16111	16072	39	0.24
7	32012	31926	86	0.26
8	64321	64110	211	0.32
9	125525	125192	333	0.26

TABLE 12 Example 5 ($d = 2$, Data 1) The strong and weak ε-accuracies of the error indicator η_{T}.

Ref.	$\#\mathcal{T}_h$ optimal	$\Theta_{strong}(\eta_T)$	$\Theta_{weak}(\eta_T, \Bbbk)$
0	200	0.00037	0
1	461	0.00112	0
2	950	0.00061	0
3	1926	0.00079	0.00518
4	3937	0.00046	0.00253
5	7963	0.00065	0.00125
6	16111	0.00041	0.00062
7	32012	0.00052	0.00010
8	64321	0.00035	0.00015
9	125525	0.00042	0.00013

Example 6. We solve the primal and dual problem pair (E_h, H_h) in adaptively refined meshes for Data 2. We compare optimal refinement achieved by using the exact error distribution e_{T} to the refinement provided by the distribution of the majorant η_{T}. Again, we start from a regular mesh with 200 elements, and perform nine refinement iterations, where on each iteration 30% of elements with the highest amount of error are refined. The converge histories are depicted in Figure 8. In Figure 9 we have depicted the meshes after the fourth refinement. Figure 10 depicts one of the finest parts of the final meshes. We see from Table 13 that the number of elements in the optimal meshes and the meshes produced using η_T are very close to each other. In fact, adaptive refinement using η_T is very close to optimal in each step, and the resulting approximation after the last refinement is practically the same. In the last mesh the combined error of the approximation pair was 0.0081176429 in the optimal mesh, and 0.0081095927 in the mesh generated by η_{T}. Table 14 shows the error indicator measures with respect to the optimal meshes. Both the strong and weak measures of the accuracy of the error indicator are very good.[3]

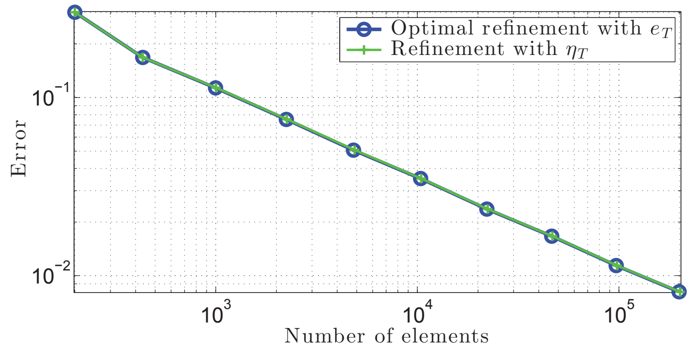


FIGURE 8 Example 6 ($d = 2$, Data 2) Adaptive computation where the error is measured in the combined norm.

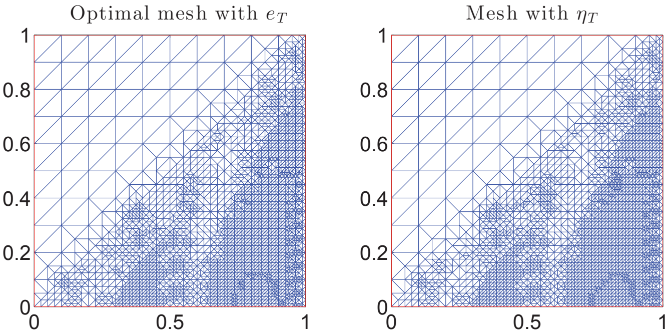


FIGURE 9 Example 6 ($d = 2$, Data 2) Adaptive mesh after the fourth refinement.

[3] This test was done also in [PIII, Example 6.6].

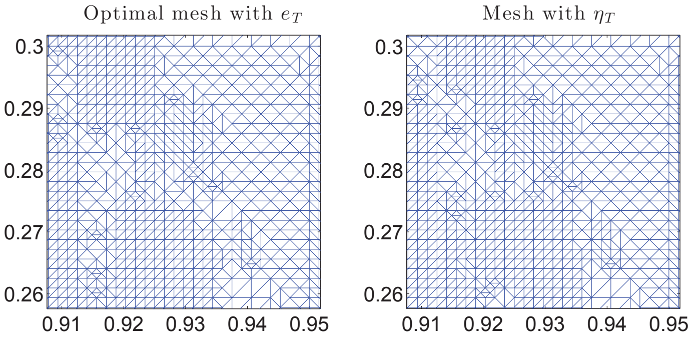


FIGURE 10 Example 6 ($d = 2$, Data 2) One of the most fine parts in the final adaptive mesh.

TABLE 13 Example 6 ($d = 2$, Data 2) The number of elements in the optimal meshes and the meshes generated by the help of η_T.

Ref.	optimal	with η_T	difference	difference %
-	200	200	0	0
1	434	434	0	0
2	998	1002	4	0.40
3	2240	2252	12	0.53
4	4823	4878	55	1.14
5	10378	10446	68	0.65
6	22116	22337	221	0.99
7	46388	46768	380	0.81
8	96859	97832	973	1.00
9	198704	200970	2266	1.14

TABLE 14 Example 6 ($d = 2$, Data 2) The strong and weak ε-accuracies of the error indicator η_{T}.

Ref.	$\#\mathcal{T}_h$ optimal	$\Theta_{strong}(\eta_T)$	$\Theta_{weak}(\eta_T, \Bbbk)$
0	200	0.0062	0
1	434	0.0120	0.00757
2	998	0.0166	0.00664
3	2240	0.0137	0.01040
4	4823	0.0143	0.00552
5	10378	0.0126	0.00545
6	22116	0.0123	0.00587
7	46388	0.0101	0.00459
8	96859	0.0105	0.00430
9	198704	0.0089	0.00394

Thus, we have established that in addition to providing the exact global error in the combined norm, it also serves as a reliable error indicator providing nearly optimal meshes in the adaptive solution process. In the following examples we will now take problems where we do not know the exact solution, i.e., we take Data 3 and 5. Since the majorant gives indeed the exact error in the combined norm, we will use this information in the following tables and figures.

Example 7. We solve the primal and dual problem pair (E_h, H_h) in adaptively refined and uniformly refined meshes for Data 3. We compare refining the whole mesh on each iteration to refining 30% of elements on each iteration using the indicator η_{T}. Again, we start from a regular mesh with 200 elements. We stop the adaptive solution process when

$$\frac{|[(E,H)-(E_h,H_h)]|}{|[E,H]|} = \frac{\mathcal{M}(E_h,H_h)^{1/2}}{\|F\|_{\kappa^{-1}}} \leq 0.007.$$

We see from Figure 11 that the adaptive procedure is beneficial in this example. From Table 15 we see that the desired accuracy was obtained on a mesh with 134205 elements. We have also depicted the approximation in Figure 12 and the mesh in Figure 13 after the fifth refinement. [4]

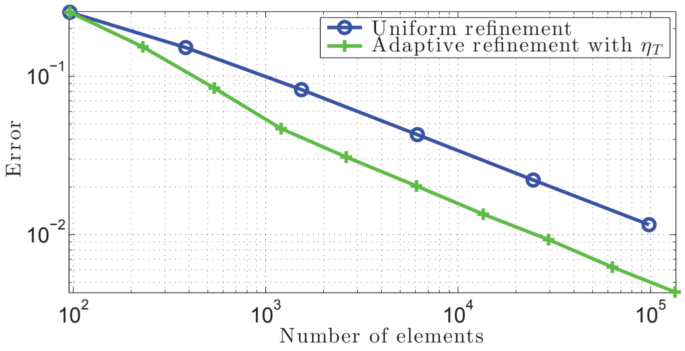


FIGURE 11 Example 7 ($d = 2$, Data 3) Adaptive computation where the error is measured in the combined norm.

[4] This test was done also in [PIII, Example 6.7]. Another similar test can be found in [PIII, Example 6.8].

TABLE 15 Example 7 ($d = 2$, Data 3) Errors in the combined norm for adaptively refined meshes.

$\#\mathcal{T}_h$	$\mathcal{M}(E_h, H_h)^{1/2}$	$\mathcal{M}(E_h, H_h)^{1/2} / \|F\|_{\kappa^{-1}}$
96	0.2534	0.2926
230	0.1534	0.1771
541	0.0842	0.0973
1204	0.0467	0.0539
2623	0.0309	0.0357
6082	0.0203	0.0234
13514	0.0135	0.0155
29530	0.0093	0.0107
63363	0.0062	0.0072
134205	0.0043	0.0050

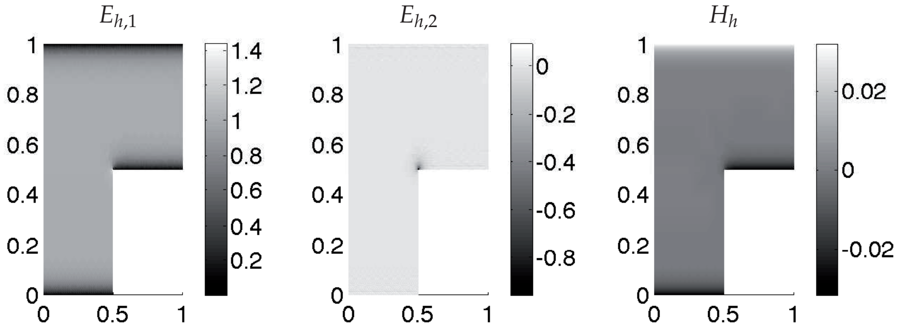


FIGURE 12 Example 7 ($d = 2$, Data 3) The two components of the approximate primal variable E_h and the dual variable H_h.

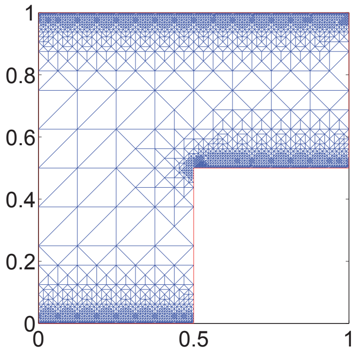


FIGURE 13 Example 7 ($d = 2$, Data 3) Adaptive mesh after the fifth refinement.

Example 8. We solve the primal and dual problem pair (E_h, H_h) in uniformly refined meshes for Data 5. We start from a regular mesh with 336 elements. We stop the solution process when

$$\frac{|[(E,H)-(E_h,H_h)]|}{|[E,H]|} = \frac{\mathcal{M}(E_h,H_h)^{1/2}}{\|F\|_{\kappa^{-1}}} \leq 0.06.$$

From Table 16 we see that the desired accuracy was obtained on a mesh after the fifth refinement with over one million elements. In Table 17 we have listed the times it took to assemble the finite element matrices of the primal problem, and how long it took to solve the linear system directly ($\mathcal{E}_h$ is the set of edges in the mesh). These runtimes do not include the time used to gather data related to the mesh (i.e., the affine transformations, etc.) These calculations were performed on a 64 processor SMP server with 1 TB of RAM. The vectorized FEM assembly routine (implemented in the way presented in [39]) starts to perform faster than the linear system solvers as the mesh size increases.

TABLE 16 Example 8 ($d = 3$, Data 5) Errors in the combined norm for uniformly refined meshes.

$\#\mathcal{T}_h$	$\mathcal{M}(E_h,H_h)^{1/2}$	$\mathcal{M}(E_h,H_h)^{1/2}/\|F\|_{\kappa^{-1}}$
42	0.3844	0.6235
336	0.2656	0.4308
2688	0.1706	0.2769
21504	0.1043	0.1691
172032	0.0619	0.1004
1376256	0.0353	0.0574

TABLE 17 Example 8 ($d = 3$, Data 5) Times (seconds) used to assemble finite element matrices for the primal variable, and to solve the linear systems directly.

$\#\mathcal{T}_h$	$\#\mathcal{E}_h$	assembly (s)	linear system (s)
42	91	0.0373	0.0006
336	548	0.0483	0.0026
2688	3736	0.168	0.030
21504	27440	2.996	0.578
172032	210016	14.883	13.598
1376256	1642688	111.336	508.290

We tested all the derived estimates and the equality. In addition to verifying the equality, we also demonstrated that the majorant serves as a nearly optimal error indicator. According to the tests made, it can be concluded that the equality provides computationally cheap both the global error, and a reliable error indicator.

4 EFFECT OF INDETERMINATE DATA

In problems related to partial differential equations it is usually assumed that the data of the problem is exactly known. However, quite often the data at hand is not complete. In many cases the data is uncertain within some intervals. Material functions, geometry, and boundary conditions may all contain uncertainty, which must be taken into account.

Studying the effects of indeterminate data gained the attention of researchers later than fully determined problems. The probabilistic approach is based in studying stochastic partial differential equations (see e.g. [45]). In [13], the authors study the so-called ”worst case scenario method”.

In this thesis we study indeterminacy by using the functional type a posteriori error equality presented in Section 3.4. First, we study the case where the right hand side F is not fully known. Then we concentrate in studying the effects of indeterminacy in the material functions μ and κ. We assume that they belong to a set of "admissible" data. This set will generate a set of solutions, and we are interested in measuring the magnitude of this set. In the last section we study the effect of indeterminate material data on error indication.

4.1 Indeterminate right hand side F

We assume that the material data (μ, κ) is known, but the right hand side F belongs to the indeterminacy set

$$\mathcal{D}_F := \{F \in L^2(\Omega) \mid \ F = F_0 + F_{osc}, \quad \|F_{osc}\|_{\kappa^{-1}} \leq \delta_F \|F_0\|_{\kappa^{-1}}\},$$

where $F_0 \in L^2(\Omega)$ denotes the known mean data, and δ_F denotes the magnitude of indeterminacy of the right hand side. We have a unique solution (E, H) for all right hand sides $F \in \mathcal{D}_F$, so the solution mapping

$$\mathcal{S} : \mathcal{D}_F \to H_{0,\Gamma_D}(\mathrm{curl}, \Omega) \times H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$$

is well defined. The solution set generated by the indeterminate data will be referred to as the set $\mathcal{S}(\mathcal{D}_F)$. We denote the mean solution generated by F_0 with

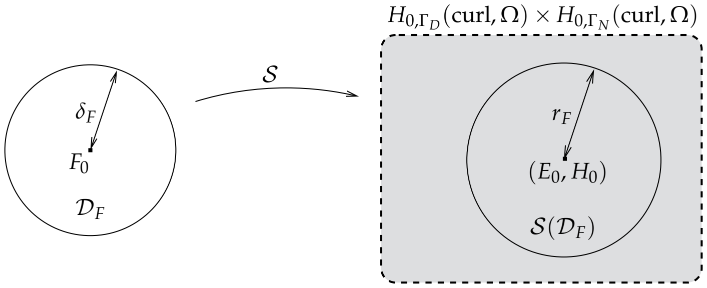

$H_{0,\Gamma_D}(\text{curl},\Omega) \times H_{0,\Gamma_N}(\text{curl},\Omega)$
$\mathcal{S}$
δ_F
F_0
$\mathcal{D}_F$
r_F
(E_0, H_0)
$\mathcal{S}(\mathcal{D}_F)$

FIGURE 14 Set of right hand sides $\mathcal{D}_F$ and the solution set $\mathcal{S}(\mathcal{D}_F)$.

(E_0, H_0). Due to the linearity of the problem all solutions can be then represented as the sum $(E, H) = (E_0, H_0) + (E_{osc}, H_{osc})$. Since by Remark 2.1 the solution operator is an isometry, we have

$$|[(E, H) - (E_0, H_0)]| = |[E_{osc}, H_{osc}]| = \|F_{osc}\|_{\kappa^{-1}} \leq \delta_F \|F_0\|_{\kappa^{-1}}.$$

Then, the distance between the mean solution and the most distant member of the solution set, the radius of the solution set, is given by

$$r_F := \sup_{(E,H)\in\mathcal{S}(\mathcal{D}_F)} |[(E, H) - (E_0, H_0)]| = \delta_F \|F_0\|_{\kappa^{-1}}, \tag{4.1}$$

and its normalized counterpart is, again by the isometry property, given by

$$\hat{r}_F := \frac{\sup_{(E,H)\in\mathcal{S}(\mathcal{D}_F)} |[(E, H) - (E_0, H_0)]|}{|[E_0, H_0]|} = \delta_F.$$

The quantities r_F and $\hat{r}_F$ are then known and fully computable. Due to the isometry property of the solution mapping the radius of the solution set is exactly the same as the magnitude of variations in the right hand side. This setting is visualized in Figure 14.

Practical application and an example

An important thing to consider is the distance of a numerical approximation to the solution set $\mathcal{S}(\mathcal{D}_F)$. Let $(\widetilde{E}, \widetilde{H}) \in H_{0,\Gamma_D}(\text{curl}, \Omega) \times H_{0,\Gamma_N}(\text{curl}, \Omega)$ be an arbitrary approximation. We define

$$\boldsymbol{\phi}_F := \frac{\sup_{(E,H)\in\mathcal{S}(\mathcal{D}_F)} |[(E, H) - (E_0, H_0)]|}{|[(E_0, H_0) - (\widetilde{E}, \widetilde{H})]|} = \frac{\delta_F \|F_0\|_{\kappa^{-1}}}{\mathcal{M}(\widetilde{E}, \widetilde{H})^{1/2}} \in [0, \infty],$$

where we have used (4.1) and the equality of Theorem 3.6. $\boldsymbol{\phi}_F$ is fully computable and gives us exact information on weather or not the approximation is inside $\mathcal{S}(\mathcal{D}_F)$.

If the value of $\boldsymbol{\phi}_F$ is close to 0, the distance to the mean solution is much bigger than the radius r_F, and we are very far from $\mathcal{S}(\mathcal{D}_F)$. The closer the value of $\boldsymbol{\phi}_F$ is to 1, the closer we are to $\mathcal{S}(\mathcal{D}_F)$. Finally, if the value is over 1, we know for sure that we are inside $\mathcal{S}(\mathcal{D}_F)$, and further computational expenditures to improve the approximation may not be useful anymore, and, i.e., adaptive processes can be stopped.

Example 9. We take Data 3 as the mean data, and assume that $\delta_F = 0.01$. We calculate the approximation (E_h, H_h) with the mean data in subsequently refined meshes untli $\boldsymbol{\phi}_F > 1$, i.e., when we are inside the solution set $\mathcal{S}(\mathcal{D}_F)$. In Table 18 are results for uniform refinement. In Table 19 we expose results for adaptive refinement using the indicator η_T defined in (3.44). On each step 30% of elements were refined. By using the adaptive procedure, the solution was inside the solution set already with 63363 elements, whereas with uniform refinement this happened with 393216 elements.

TABLE 18 Example 9 ($d = 2$, Data 3) Values of $\boldsymbol{\phi}_F$ in the case of uniform refinement.

$\#\mathcal{T}_h$	$\mathcal{M}(E_h, H_h)^{1/2}$	$\mathcal{M}(E_h, H_h)^{1/2} / \|F_0\|_{\kappa^{-1}}$	$\boldsymbol{\phi}_F$
96	0.25344	0.29265	0.034
384	0.15183	0.17532	0.057
1536	0.08235	0.09508	0.105
6144	0.04282	0.04945	0.202
24576	0.02214	0.02556	0.391
98304	0.01155	0.01333	0.749
393216	0.00612	0.00707	1.413

TABLE 19 Example 9 ($d = 2$, Data 3) Values of $\boldsymbol{\phi}_F$ in the case of adaptive refinement.

$\#\mathcal{T}_h$	$\mathcal{M}(E_h, H_h)^{1/2}$	$\mathcal{M}(E_h, H_h)^{1/2} / \|F_0\|_{\kappa^{-1}}$	$\boldsymbol{\phi}_F$
96	0.25344	0.29265	0.034
230	0.15336	0.17709	0.056
541	0.08422	0.09725	0.102
1204	0.04667	0.05389	0.185
2623	0.03089	0.03567	0.280
6082	0.02029	0.02343	0.426
13514	0.01345	0.01554	0.643
29530	0.00930	0.01074	0.930
63363	0.00623	0.00720	1.388

4.2 Indeterminate material data μ and κ

Two-sided estimates of the radius of the solution set in the case of indeterminate material data were first studied in [19–21] in the context of reaction-diffusion type problems. In the extended abstract [1] a similar result for a magnetostatics problem was presented. In this section we use same techniques to derive estimates of the radius of the solution set for the model problem considered in this thesis.

We assume that the right hand side F is known, but the material functions μ and κ belong to the indeterminacy sets

$$\mathcal{D}_1 := \{\mu \in L^\infty(\overline{\Omega}) \mid \mu = \mu_0 + \varphi_1,\ -\delta_1 \le \varphi_1(x) \le \delta_1 \text{ for a.e. } x \in \overline{\Omega}\},$$
$$\mathcal{D}_2 := \{\kappa \in L^\infty(\overline{\Omega}) \mid \kappa = \kappa_0 + \varphi_2,\ -\delta_2 \le \varphi_2(x) \le \delta_2 \text{ for a.e. } x \in \overline{\Omega}\},$$

where $\mu_0, \kappa_0 \in L^\infty(\overline{\Omega})$ is the mean data, and δ_1 and δ_2 are magnitudes of variation in the material data. We assume that

$$0 < \underline{c}_1 \le \mu_0(x) \le \overline{c}_1 < \infty,$$
$$0 < \underline{c}_2 \le \kappa_0(x) \le \overline{c}_2 < \infty,$$

for a.e. $x \in \overline{\Omega}$. With

$$0 \le \delta_1 < \underline{c}_1, \tag{4.2}$$
$$0 \le \delta_2 < \underline{c}_2, \tag{4.3}$$

we have a unique solution (E, H) for all $(\mu, \kappa) \in \mathcal{D} := \mathcal{D}_1 \times \mathcal{D}_2$, and the solution mapping $\mathcal{S} : \mathcal{D} \to H_{0,\Gamma_D}(\text{curl}, \Omega) \times H_{0,\Gamma_N}(\text{curl}, \Omega)$ is well defined. The solution set generated by the indeterminate data will be referred to as the set $\mathcal{S}(\mathcal{D})$. Each pair of material functions $(\mu, \kappa) \in \mathcal{D}$ generate the norms

$$\begin{aligned} |\!|\!| v |\!|\!|^2_{\triangleright} &:= a_{\triangleright}(v, v) := \|\text{curl}\, v\|^2_{\mu^{-1}} + \|v\|^2_{\kappa}, \\ |\!|\![q]\!|\!|^2_{\triangleright} &:= \hat{a}_{\triangleright}(q, q) := \|\text{curl}\, q\|^2_{\kappa^{-1}} + \|q\|^2_{\mu}, \\ |[v, q]|^2_{\triangleright} &:= |\!|\!| v |\!|\!|^2_{\triangleright} + |\!|\![q]\!|\!|^2_{\triangleright}\,. \end{aligned}$$

The norms generated by the mean data (μ_0, κ_0) are denoted by $|\!|\!| \cdot |\!|\!|_0$, $|\!|\![\cdot]\!|\!|_0$ and $|[\cdot]|_0$. The mean solution generated by this data is denoted by (E_0, H_0). Also in the case of other quantities, we will always associate the subindex $\triangleright$ to the indeterminate data, and the subindex 0 to the mean data.

Two important quantities governing the radius of the solution set are the ratios between the magnitudes of variation and the lowest values of the mean data:

$$\theta_1 := \frac{\delta_1}{\underline{c}_1} \qquad \text{and} \qquad \theta_2 := \frac{\delta_2}{\underline{c}_2}.$$

It is clear that $0 \le \theta_1, \theta_2 < 1$. Also, we define the ratios between the largest and lowest values of mean data:

$$\gamma_1 := \frac{\overline{c}_1}{\underline{c}_1} \qquad \text{and} \qquad \gamma_2 := \frac{\overline{c}_2}{\underline{c}_2}.$$

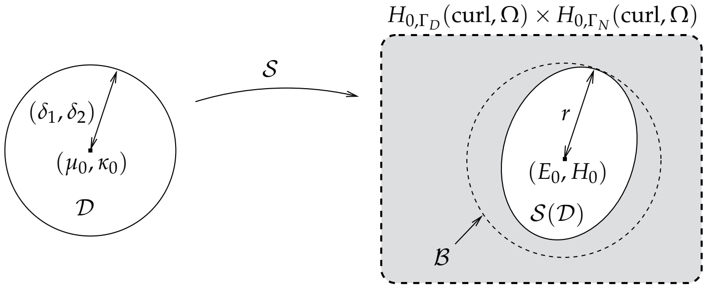

$H_{0,\Gamma_D}(\mathrm{curl},\Omega) \times H_{0,\Gamma_N}(\mathrm{curl},\Omega)$
$\mathcal{S}$
(δ_1, δ_2)
(μ_0, κ_0)
$\mathcal{D}$
r
(E_0, H_0)
$\mathcal{S}(\mathcal{D})$
$\mathcal{B}$

FIGURE 15 Set of material data $\mathcal{D}$ and the solution set $\mathcal{S}(\mathcal{D})$.

We have $\gamma_1, \gamma_2 \geq 1$.

We are interested in the distance between the mean solution and the most distant member of the solution set. We call this quantity the radius of the solution set, and we measure it in the combined norm:

$$r := \sup_{(E,H)\in\mathcal{S}(\mathcal{D})} |[(E,H) - (E_0,H_0)]|_0 \,. \tag{4.4}$$

The mean solution and the radius define the ball $\mathcal{B} := \overline{B}((E_0,H_0),r)$ in the sense of the mean combined norm. It is clear that this ball contains the solution set: $\mathcal{S}(\mathcal{D}) \subset \mathcal{B}$. This situation is illustrated in Figure 15.

Proposition 4.1. *Under the above made assumptions, for any v, q we have*

$$\underline{K}_E \, \|\!| v \|\!|_{\triangleright}^2 \;\leq\; \|\!| v \|\!|_0^2 \;\leq\; \overline{K}_E \, \|\!| v \|\!|_{\triangleright}^2, \tag{4.5}$$

$$\underline{K}_H \, |\!\|\!| q |\!\|\!|_{\triangleright}^2 \;\leq\; |\!\|\!| q |\!\|\!|_0^2 \;\leq\; \overline{K}_H \, |\!\|\!| q |\!\|\!|_{\triangleright}^2, \tag{4.6}$$

$$\underline{K} \, |[v,q]|_{\triangleright}^2 \;\leq\; |[v,q]|_0^2 \;\leq\; \overline{K} \, |[v,q]|_{\triangleright}^2, \tag{4.7}$$

where

$$\underline{K}_E := \min\left(1-\theta_1, \frac{1}{1+\theta_2}\right), \qquad \overline{K}_E := \max\left(1+\theta_1, \frac{1}{1-\theta_2}\right),$$

$$\underline{K}_H := \min\left(1-\theta_2, \frac{1}{1+\theta_1}\right), \qquad \overline{K}_H := \max\left(1+\theta_2, \frac{1}{1-\theta_1}\right),$$

$$\underline{K} := \min\left(1-\theta_1, 1-\theta_2\right), \qquad \overline{K} := \max\left(\frac{1}{1-\theta_1}, \frac{1}{1-\theta_2}\right).$$

Proof. First we note that

$$\mu_0^{-1}\left(\frac{1}{1+\theta_1}\right) = \frac{1}{\mu_0}\left(\frac{1}{1+\frac{\delta_1}{\underline{c}_1}}\right) \leq \frac{1}{\mu_0}\left(\frac{1}{1+\frac{\delta_1}{\mu_0}}\right) = \frac{1}{\mu_0}\left(\frac{\mu_0}{\mu_0+\delta_1}\right) \leq$$

$$\leq \; \mu^{-1} = \frac{1}{\mu_0+\varphi} = \frac{1}{\mu_0}\left(\frac{\mu_0}{\mu_0+\varphi}\right) \; \leq$$

$$\leq \frac{1}{\mu_0}\left(\frac{\mu_0}{\mu_0-\delta_1}\right) = \frac{1}{\mu_0}\left(\frac{1}{1-\frac{\delta_1}{\mu_0}}\right) \leq \frac{1}{\mu_0}\left(\frac{1}{1-\frac{\delta_1}{\underline{c}_1}}\right) = \mu_0^{-1}\left(\frac{1}{1-\theta_1}\right), \tag{4.8}$$

and similarly

$$\begin{aligned}\kappa_0\,(1-\theta_2) = \kappa_0\left(1-\frac{\delta_2}{\underline{c}_2}\right) \le \kappa_0\left(1-\frac{\delta_2}{\kappa_0}\right) \le \\ \le\ \kappa = \kappa_0 + \phi_2 = \kappa_0\left(1+\frac{\phi_2}{\kappa_0}\right)\ \le \\ \le \kappa_0\left(1+\frac{\delta_2}{\kappa_0}\right) \le \kappa_0\left(1+\frac{\delta_2}{\underline{c}_2}\right) = \kappa_0\,(1+\theta_2)\,. \quad (4.9)\end{aligned}$$

With (4.8) and (4.9) we have

$$\min\left(\frac{1}{1+\theta_1}, 1-\theta_2\right)\ |\!|\!|\, v\, |\!|\!|_0^2 \le |\!|\!|\, v\, |\!|\!|_\triangleright^2 \le \max\left(\frac{1}{1-\theta_1}, 1+\theta_2\right)\ |\!|\!|\, v\, |\!|\!|_0^2,$$

which implies (4.5). By performing the calculation (4.8) for κ^{-1}, and the calculation (4.9) for μ, we obtain

$$\min\left(\frac{1}{1+\theta_2}, 1-\theta_1\right)\ |\!|\!|[\, q\,]|\!|\!|_0^2 \le |\!|\!|[\, q\,]|\!|\!|_\triangleright^2 \le \max\left(\frac{1}{1-\theta_2}, 1+\theta_1\right)\ |\!|\!|[\, q\,]|\!|\!|_0^2,$$

which implies (4.6). The final norm equivalence (4.7) is a direct consequence of (4.5) and (4.6).

□

By using the equality of Theorem 3.6 we can now estimate the radius (4.4) from above by

$$r^2\ \le \sup_{(E,H)\in\mathcal{S}(\mathcal{D})} \overline{K}\,\|[(E,H)-(E_0,H_0)]\|_\triangleright^2\ =\ \overline{K}\sup_{(\mu,\kappa)\in\mathcal{D}} \mathcal{M}_\triangleright(E_0,H_0) \qquad (4.10)$$

and from below by

$$r^2\ \ge \sup_{(E,H)\in\mathcal{S}(\mathcal{D})} \underline{K}\,\|[(E,H)-(E_0,H_0)]\|_\triangleright^2\ =\ \underline{K}\sup_{(\mu,\kappa)\in\mathcal{D}} \mathcal{M}_\triangleright(E_0,H_0). \qquad (4.11)$$

Theorem 4.1. *Let E_0 be the exact mean electric field. The radius of the solution set is subject to the two-sided estimate*

$$r^2\ \le\ \overline{K}\left[\left(\frac{\theta_2^2}{1-\theta_2}\right)\|E_0\|_{\kappa_0}^2 + \left(\frac{\theta_1^2}{1-\theta_1}\right)\|\mathrm{curl}\,E_0\|_{\mu_0^{-1}}^2\right], \qquad (4.12)$$

$$r^2\ \ge\ \underline{K}\left[\left(\frac{\theta_2^2}{\gamma_2(\gamma_2-\theta_2)}\right)\|E_0\|_{\kappa_0}^2 + \left(\frac{\theta_1^2}{\gamma_1(\gamma_1-\theta_1)}\right)\|\mathrm{curl}\,E_0\|_{\mu_0^{-1}}^2\right]. \qquad (4.13)$$

Proof. In order to estimate (4.10) and (4.11) we first note that

$$\begin{aligned}\sup_{(\mu,\kappa)\in\mathcal{D}} \mathcal{M}_\triangleright(E_0,H_0) &= \sup_{(\mu,\kappa)\in\mathcal{D}}\left[\|F-\kappa E_0-\mathrm{curl}\,H_0\|_{\kappa^{-1}}^2 + \|H_0-\mu^{-1}\mathrm{curl}\,E_0\|_\mu^2\right] \\ &= \sup_{(\mu,\kappa)\in\mathcal{D}}\left[\|\kappa_0E_0-\kappa E_0\|_{\kappa^{-1}}^2 + \|\mu_0^{-1}\mathrm{curl}\,E_0-\mu^{-1}\mathrm{curl}\,E_0\|_\mu^2\right] \\ &= \sup_{\kappa\in\mathcal{D}_2}\int_\Omega \kappa^{-1}(\kappa_0-\kappa)^2\,|E_0|^2\,\mathrm{dx}\ + \\ &\quad + \sup_{\mu\in\mathcal{D}_1}\int_\Omega \mu(\mu_0^{-1}-\mu^{-1})^2\,|\mathrm{curl}\,E_0|^2\,\mathrm{dx}. \qquad (4.14)\end{aligned}$$

Since

$$\kappa^{-1}(\kappa_0-\kappa)^2=\frac{\varphi_2^2}{\kappa_0+\varphi_2}=\kappa_0\left(\frac{\varphi_2^2}{\kappa_0(\kappa_0+\varphi_2)}\right)$$

and

$$\mu(\mu_0^{-1}-\mu^{-1})^2=(\mu_0+\varphi_1)\left(\frac{1}{\mu_0}-\frac{1}{\mu_0+\varphi_1}\right)^2=$$
$$=(\mu_0+\varphi_1)\left(\frac{\varphi_1}{\mu_0(\mu_0+\varphi_1)}\right)^2=\frac{\varphi_1^2}{\mu_0^2(\mu_0+\varphi_1)}=\mu_0^{-1}\left(\frac{\varphi_1^2}{\mu_0(\mu_0+\varphi_1)}\right),$$

we can write (4.14) as

$$\sup_{(\mu,\kappa)\in\mathcal{D}}\mathcal{M}_{\triangleright}(E_0,H_0)=\int_\Omega\kappa_0\left(\frac{\delta_2^2}{\kappa_0(\kappa_0-\delta_2)}\right)|E_0|^2\,\mathrm{dx}+$$
$$+\int_\Omega\mu_0^{-1}\left(\frac{\delta_1^2}{\mu_0(\mu_0-\delta_1)}\right)|\mathrm{curl}\,E_0|^2\,\mathrm{dx}.\quad(4.15)$$

Then we estimate (4.15) from above by

$$\sup_{(\mu,\kappa)\in\mathcal{D}}\mathcal{M}_{\triangleright}(E_0,H_0)\le\left(\frac{\delta_2^2}{\underline{c}_2(\underline{c}_2-\delta_2)}\right)\|E_0\|_{\kappa_0}^2+\left(\frac{\delta_1^2}{\underline{c}_1(\underline{c}_1-\delta_1)}\right)\|\mathrm{curl}\,E_0\|_{\mu_0^{-1}}^2$$

and we have proven (4.12). Similarly, we estimate (4.15) from below by

$$\sup_{(\mu,\kappa)\in\mathcal{D}}\mathcal{M}_{\triangleright}(E_0,H_0)\ge\left(\frac{\delta_2^2}{\overline{c}_2(\overline{c}_2-\delta_2)}\right)\|E_0\|_{\kappa_0}^2+\left(\frac{\delta_1^2}{\overline{c}_1(\overline{c}_1-\delta_1)}\right)\|\mathrm{curl}\,E_0\|_{\mu_0^{-1}}^2$$

and we have proven (4.13).
□

Theorem 4.2. *Let E_0 be the exact mean electric field, and $\delta_1,\delta_2>0$. We then have the estimate*

$$\underline{R}\,|||\,E_0\,|||_0^2\;\le\;r^2\le\;\overline{R}\,|||\,E_0\,|||_0^2,\quad(4.16)$$

where

$$\underline{R}:=\underline{K}\min\left(\frac{\theta_1^2}{\gamma_1(\gamma_1-\theta_1)},\frac{\theta_2^2}{\gamma_2(\gamma_2-\theta_2)}\right),\quad(4.17)$$

$$\overline{R}:=\max\left(\frac{\theta_1}{1-\theta_1},\frac{\theta_2}{1-\theta_2}\right)^2.\quad(4.18)$$

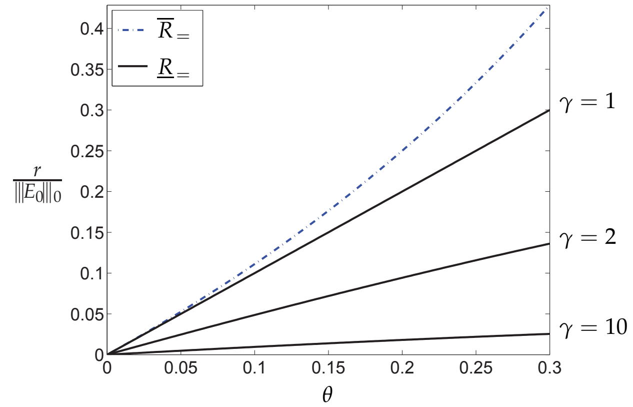


FIGURE 16 The two-sided estimate (4.19) of the normalized radius of the solution set $\frac{r}{|\!|\!| E_0 |\!|\!|_0}$ with different γ.

Proof. Without additional assumptions we can estimate the upper bound (4.12) further from above by

$$r^2 \leq \overline{K} \max\left(\frac{\theta_1^2}{1-\theta_1}, \frac{\theta_2^2}{1-\theta_2}\right) |\!|\!| E_0 |\!|\!|_0^2 = \max\left(\frac{\theta_1}{1-\theta_1}, \frac{\theta_2}{1-\theta_2}\right)^2 |\!|\!| E_0 |\!|\!|_0^2 .$$

By assuming that $\delta_1, \delta_2 > 0$ the lower bound (4.13) can be estimated by

$$r^2 \geq \underline{K} \min\left(\frac{\theta_1^2}{\gamma_1(\gamma_1-\theta_1)}, \frac{\theta_2^2}{\gamma_2(\gamma_2-\theta_2)}\right) |\!|\!| E_0 |\!|\!|_0^2 .$$

□

Let us assume that material data is indeterminate such that $\delta := \delta_1 = \delta_2 > 0$, $\overline{c} := \overline{c}_1 = \overline{c}_2$, and $\underline{c} := \underline{c}_1 = \underline{c}_2$. Then, by Theorem 4.2, and by normalizing the radius r by $|\!|\!| E_0 |\!|\!|_0$ we see that

$$\underline{R}_= := \left(\frac{(1-\theta)\theta^2}{\gamma(\gamma-\theta)}\right)^{1/2} \leq \frac{r}{|\!|\!| E_0 |\!|\!|_0} \leq \frac{\theta}{1-\theta} =: \overline{R}_=, \tag{4.19}$$

where $\theta := \frac{\delta}{\underline{c}}$ and $\gamma := \frac{\overline{c}}{\underline{c}}$. In Figure 16 we have visualized the bounds $\overline{R}_=$ and $\underline{R}_=$. Since the lower bound depends on the ratio of the biggest and smallest values of the mean data, it becomes very small as this ratio grows.

Practical application and an example

An important thing to consider is the distance of a numerical approximation to the solution set $\mathcal{S}(\mathcal{D})$. However, it is easier to obtain computable quantities with respect to the ball $\mathcal{B}$ which contains the solution set.

Let $(\widetilde{E}, \widetilde{H}) \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega) \times H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$ be an arbitrary approximation. Our quantity of interest is

$$\boldsymbol{\phi} := \frac{\sup_{(E,H)\in\mathcal{S}(\mathcal{D})} |[(E,H) - (E_0,H_0)]|_0}{|[(E_0,H_0) - (\widetilde{E},\widetilde{H})]|_0} = \frac{r}{\mathcal{M}_0(\widetilde{E},\widetilde{H})^{1/2}} \in [0,\infty). \tag{4.20}$$

Here we again used the equality of Theorem 3.6. If the value of $\boldsymbol{\phi}$ is close to 0, the distance to the mean solution is much bigger than the radius r, and we are very far from $\mathcal{B}$. The closer the value of $\boldsymbol{\phi}$ is to 1, the closer we are to $\mathcal{B}$. Finally, if the value is over 1, we know for sure that we are inside $\mathcal{B}$, and we are in the range of $\mathcal{S}(\mathcal{D})$. The greater the value of $\boldsymbol{\phi}$, the closer we are to the mean solution (E_0, H_0), and the bigger the probability that the approximation is inside the $\mathcal{S}(\mathcal{D})$. The situation illustrated in Figure 17.

The quantity $\boldsymbol{\phi}$ is not directly computable. However, we are able to derive the following computable bounds for this quantity.

Theorem 4.3. *Let $\delta_1, \delta_2 > 0$. For any $(\widetilde{E}, \widetilde{H}) \in H_{0,\Gamma_D}(\mathrm{curl}, \Omega) \times H_{0,\Gamma_N}(\mathrm{curl}, \Omega)$ we have*

$$\boldsymbol{\phi}_{\ominus}(\widetilde{E},\widetilde{H}) \;\le\; \boldsymbol{\phi} \;\le\; \boldsymbol{\phi}_{\oplus}(\widetilde{E},\widetilde{H}),$$

where

$$\boldsymbol{\phi}_{\ominus}(\widetilde{E},\widetilde{H}) := \left(\underline{R}\frac{-2\mathcal{J}_0(\widetilde{E})}{\mathcal{M}_0(\widetilde{E},\widetilde{H})}\right)^{1/2}, \qquad \boldsymbol{\phi}_{\oplus}(\widetilde{E},\widetilde{H}) := \overline{R}^{1/2}\frac{|\!|\!| \widetilde{E} |\!|\!|_0 + \mathcal{M}_0(\widetilde{E},\widetilde{H})^{1/2}}{\mathcal{M}_0(\widetilde{E},\widetilde{H})^{1/2}}.$$

The constants $\underline{R}$ and $\overline{R}$ are defined in (4.17)–(4.18)*, $\mathcal{J}_0$ is the energy functional* (2.34) *with the mean data, and $\mathcal{M}_0$ is the majorant defined in Theorem 3.6 with the mean data.*

Proof. By using the Minkowski inequality (2.5) we have

$$|\!|\!| E_0 |\!|\!|_0 = |\!|\!| E_0 - \widetilde{E} + \widetilde{E} |\!|\!|_0 = |\!|\!| \widetilde{E} |\!|\!|_0 + |\!|\!| E_0 - \widetilde{E} |\!|\!|_0 \le |\!|\!| \widetilde{E} |\!|\!|_0 + \mathcal{M}_0(\widetilde{E},\widetilde{H})^{1/2},$$

and together with (4.16) and (4.20) we have the upper bound $\boldsymbol{\phi}_{\oplus}$. To prove the lower bound we first note that

$$\mathcal{J}_0(E_0) = -\frac{1}{2} |\!|\!| E_0 |\!|\!|_0^2 .$$

Then, since by definition $\mathcal{J}_0(E_0) \le \mathcal{J}_0(\widetilde{E})$, we have

$$|\!|\!| E_0 |\!|\!|^2 \ge -2\mathcal{J}_0(\widetilde{E}),$$

and together with (4.16) and (4.20) we have the lower bound $\boldsymbol{\phi}_{\ominus}$.
□

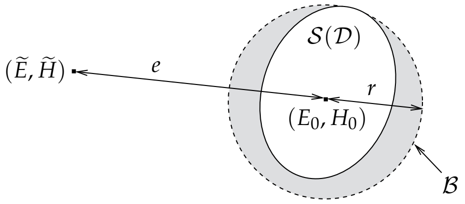


FIGURE 17 The distance to the mean solution $e = \|[(E_0, H_0) - (\widetilde{E}, \widetilde{H})]\|$ versus the radius of the solution set r.

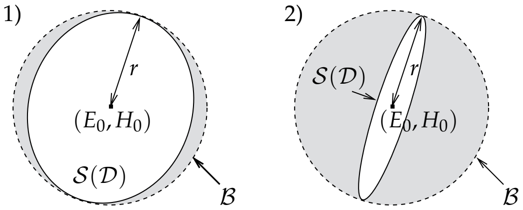


FIGURE 18 Configurations of the solution set $\mathcal{S}(\mathcal{D})$ inside the ball $\mathcal{B}$.

The bounds $\boldsymbol{\phi}_\ominus$ and $\boldsymbol{\phi}_\oplus$ are fully computable: they contain only the problem data and a conforming numerical approximation $(\widetilde{E}, \widetilde{H})$. No additional computations are needed. These bounds can be used to estimate weather or not the approximation $(\widetilde{E}, \widetilde{H})$ is close enough, or inside the ball $\mathcal{B}$. If $\boldsymbol{\phi}_\oplus > 1$, there is a possibility that the solution is inside $\mathcal{B}$. Improving the approximation should be continued *at least* to this point. However, to be sure that the approximation is inside $\mathcal{B}$, we must have $\boldsymbol{\phi}_\ominus > 1$.

However, it should be noted that even when the approximation is inside $\mathcal{B}$, the approximation may still be relatively far from the solution set $\mathcal{S}(\mathcal{D})$. This depends greatly on the topology of the solution set (see Figure 18).

Example 10. We test the bounds of Theorem 4.3. We take Data 3 as the mean data, and assume that $\delta_1 = 100$ and $\delta_2 = 0.1$. Then $\theta_1 = \theta_2 = 0.1$. Note than we also have $\gamma_1 = \gamma_2 = 1$, so the constants $\underline{R}$ and $\overline{R}$ are relatively close to each other. We calculate the approximation (E_h, H_h) with the mean data in subsequently refined meshes untli $\boldsymbol{\phi}_\ominus > 1$, i.e., when we are sure that the approximation is inside the ball $\mathcal{B}$. In Table 20 are results for uniform refinement. In Table 21 we expose results for adaptive refinement using the indicator η_{T} defined in (3.44). On each step 30% of elements were refined. By using the adaptive procedure, the solution was inside the solution set already with 63363 elements, whereas with uniform refinement this happened with 393216 elements.

TABLE 20 Example 10 ($d = 2$, Data 3) Bounds of $\boldsymbol{\phi}$ in the case of uniform refinement.

$\#\mathcal{T}_h$	$\mathcal{M}_0(E_h, H_h)^{1/2}$	$\mathcal{M}_0(E_h, H_h)^{1/2} / \|F\|_{\kappa_0^{-1}}$	$\boldsymbol{\phi}_\ominus$	$\boldsymbol{\phi}_\oplus$
96	0.25344	0.29265	0.159	0.288
384	0.15183	0.17532	0.210	0.344
1536	0.08235	0.09508	0.287	0.431
6144	0.04282	0.04945	0.400	0.555
24576	0.02214	0.02556	0.556	0.729
98304	0.01155	0.01333	0.771	0.968
393216	0.00612	0.00707	1.059	1.287

TABLE 21 Example 10 ($d = 2$, Data 3) Bounds of $\boldsymbol{\phi}$ in the case of adaptive refinement.

$\#\mathcal{T}_h$	$\mathcal{M}_0(E_h, H_h)^{1/2}$	$\mathcal{M}_0(E_h, H_h)^{1/2} / \|F\|_{\kappa_0^{-1}}$	$\boldsymbol{\phi}_\ominus$	$\boldsymbol{\phi}_\oplus$
96	0.25344	0.29265	0.159	0.288
230	0.15336	0.17709	0.209	0.343
541	0.08422	0.09725	0.284	0.427
1204	0.04667	0.05389	0.383	0.537
2623	0.03089	0.03567	0.471	0.634
6082	0.02029	0.02343	0.581	0.757
13514	0.01345	0.01554	0.714	0.904
29530	0.00930	0.01074	0.859	1.066
63363	0.00623	0.00720	1.049	1.277

4.3 Indeterminate material data and error indication

Studying the effects of indeterminate data in the context of error indication is a very relevant question since mesh adaptive numerical methods are widely used. To the best of our knowledge, the first study into this direction was done in the included conference article [PV]. In this article we studied the two-dimensional diffusion problem with indeterminate diffusion matrix. In this section we will repeat this analysis to the eddy-current problem. We adapt the notation used in Section 4.2.

We assume that we calculate our approximations and error indicators in a regular mesh $\mathcal{T}_h$. In our analysis, we consider small disturbances of the material data of the form

$$\mu = \mu_0 + \delta_1 \mu^{\pm},$$
$$\kappa = \kappa_0 + \delta_2 \kappa^{\pm},$$

where the magnitude of variations δ_1 and δ_2 satisfiy the conditions given in (4.2)–(4.3). For each element T, the elements of $\mu^{\pm}$ and $\kappa^{\pm}$ are chosen as follows:

$$\mu^{\pm}\big|_{\mathrm{T}} \in \{-1, 0, 1\}, \qquad \kappa^{\pm}\big|_{\mathrm{T}} \in \{-1, 0, 1\}, \qquad \forall \mathrm{T} \in \mathcal{T}_h.$$

In other words, we generate a constant perturbation of magnitudes δ_1 and δ_2 in each element T. In our computations we choose a very particular type of perturbation where each element T having a value 1 is surrounded by elements having values -1. This sign distribution is illustrated in the Figure 19 for $d = 2$. We denote this distribution by φ. A perturbation generated in this way is clearly an extreme one. It suits our purposes, since we are trying to find a worst case situation that can occur with different material data from the indeterminacy set $\mathcal{D}$. A set of 8 perturbed material data (μ_i, κ_i) is displayed in Table 22. We also generated another set of perturbations using a distribution φ_0 which is otherwise same as φ, but it contains 0 for those elements, which were marked to be refined by the indicator calculated with the mean data (μ_0, κ_0). So the total number of perturbed material data used in the computations is 16. In the case where only one of the material parameters contain indeterminacy, we obtain with the same principle 4 different perturbations.

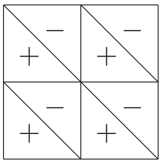


FIGURE 19 An extreme sign distribution for a regular mesh when $d = 2$.

TABLE 22 Different material data (μ_i, κ_i) used in computations.

j	1	2	3	4	5	6	7	8
$\mu^{\pm}$	$+\varphi$	$-\varphi$	0	0	$+\varphi$	$+\varphi$	$-\varphi$	$-\varphi$
$\kappa^{\pm}$	0	0	$+\varphi$	$-\varphi$	$+\varphi$	$-\varphi$	$+\varphi$	$-\varphi$

We will test the indicator η_{T} motivated by the error equality of Theorem 3.6 and defined in (3.44). We test η_{T} with respect to the marker $\Bbbk$ which refines 30% of the elements which contain the highet amount of error.

Our analysis of effects caused by data indeterminacy is based on the following method. We select a mesh $\mathcal{T}_h$ and select the functions (μ_i, κ_i) (generated in the way described before) for some given δ_1 and δ_2. For each exact solution $(E_i, H_i) = \mathcal{S}(\mu_i, \kappa_i)$, we compute the corresponding approximations $(E_{i,h}, H_{i,h})$ on the mesh $\mathcal{T}_h$. Then, for each $(E_{i,h}, H_{i,h})$, we calculate the error indicator

$$\eta_{\mathrm{T},i} := \mathcal{M}(E_{i,h}, H_{i,h})\big|_{\mathrm{T}} := \|F - \kappa_i E_{i,h} - \operatorname{curl} H_{i,h}\|^2_{\mathrm{T},\kappa_i^{-1}} + \|H_{i,h} - \mu_i^{-1} \operatorname{curl} E_{i,h}\|^2_{\mathrm{T},\mu_i},$$

and the corresponding markings $\Bbbk(\eta_{\mathrm{T},i})$.

The difference of two markings is given by the boolean measure

$$\Theta_{pair}(\Bbbk, \eta_{\mathrm{T},i}, \eta_{\mathrm{T},j}) \ := \ 1 - \frac{\sum(\Bbbk(\eta_{\mathrm{T},i}) \wedge \Bbbk(\eta_{\mathrm{T},j}))}{0.3 \cdot \#\mathcal{T}_h} \ \in \ [0, 1],$$

where $\wedge$ is the logical `AND` operator. If $\Theta_{ind}(\Bbbk, \eta_{\mathrm{T},i}, \eta_{\mathrm{T},j}) = 0$, then variations of the data do not affect the process of marking. In opposite, if $\Theta_{ind}(\Bbbk, \eta_{\mathrm{T},i}, \eta_{\mathrm{T},j})$ is close to one, then the lists of elements selected for refinement by $\eta_{\mathrm{T},i}$ and $\eta_{\mathrm{T},j}$ are quite different. The maximal difference between all markings is given by the quantity

$$\Theta_{max} \ := \ \max_{i,j}\{\Theta_{pair}(\Bbbk, \eta_{\mathrm{T},i}, \eta_{\mathrm{T},j})\},$$

which shows the maximal difference produced by an error indicator with different material data from the set $\mathcal{D}$ generated in the way described earlier in this section.

In the numerical results we will choose the magnitudes of variation δ_1 and δ_2 in such a way that their normalized counterparts

$$\theta_1 = \frac{\delta_1}{\underline{c}_1} \qquad \text{and} \qquad \theta_2 = \frac{\delta_2}{\underline{c}_2}$$

obtain values from the set $\{0.005, 0.01, 0.02, 0.03, 0.04, 0.05, 0.1\}$.

Example 11. We test the effect of indeterminacy for Data 1. In Table 23 and Table 24 are results where μ and κ contain indeterminacy separately. In Table 25 are results where both material parameters contain indeterminacy simultaneously. Our choice of perturbations result in very small values of Θ_{max} if only μ is assumed to be indeterminate. If κ, or both of the material parameters are indeterminate, then error indication starts to get affected. However, for this particular data, even the highest values of Θ_{max} are relatively small.

TABLE 23 Example 11 ($d = 2$, Data 1) Values of Θ_{max} for indeterminate μ.

	$\theta_1 > 0$ $(\theta_2 = 0)$						
$\#\mathcal{T}_h$	0.005	0.01	0.02	0.03	0.04	0.05	0.1
800	0.00414	0.00414	0.00414	0.00414	0.00414	0.00414	0.00414
3200	0.00104	0.00104	0.00104	0.00104	0.00104	0.00104	0.00312
12800	0.00026	0.00026	0.00078	0.00078	0.00078	0.00182	0.00182
51200	0.00006	0.00019	0.00032	0.00032	0.00045	0.00058	0.00149
115200	0.00002	0.00014	0.00026	0.00037	0.00054	0.00066	0.00164

TABLE 24 Example 11 ($d = 2$, Data 1) Values of Θ_{max} for indeterminate κ.

	$\theta_2 > 0$ $(\theta_1 = 0)$						
$\#\mathcal{T}_h$	0.005	0.01	0.02	0.03	0.04	0.05	0.1
800	0.0041	0.0124	0.0290	0.0456	0.0580	0.0705	0.1203
3200	0.0052	0.0093	0.0124	0.0218	0.0280	0.0447	0.0884
12800	0.0070	0.0137	0.0242	0.0356	0.0439	0.0518	0.1116
51200	0.0099	0.0188	0.0380	0.0572	0.0772	0.0964	0.1934
115200	0.0134	0.0274	0.0563	0.0843	0.1117	0.1405	0.2861

TABLE 25 Example 11 ($d = 2$, Data 1) Values of Θ_{max} for indeterminate μ and κ.

	$\theta_1 = \theta_2 > 0$						
$\#\mathcal{T}_h$	0.005	0.01	0.02	0.03	0.04	0.05	0.1
800	0.0041	0.0124	0.0290	0.0456	0.0580	0.0705	0.1203
3200	0.0052	0.0093	0.0124	0.0239	0.0280	0.0468	0.0884
12800	0.0070	0.0137	0.0247	0.0356	0.0445	0.0528	0.1116
51200	0.0102	0.0190	0.0382	0.0576	0.0774	0.0964	0.1934
115200	0.0135	0.0274	0.0563	0.0844	0.1119	0.1408	0.2861

The effect of indeterminate μ was insignificant for other test data as well, so those results are excluded. In the following tests we will report only the effect of indeterminate κ.

Example 12. We test the effect of indeterminate κ. Results for Data 2, 3, and 5 are in Tables 26, 27, and 28, respectively. For these examples indeterminacy of κ results in very big values of Θ_{max} for larger meshes.

TABLE 26 Example 12 ($d = 2$, Data 2) Values of Θ_{max} for indeterminate κ.

	$\theta_2 > 0$ $(\theta_1 = 0)$						
$\#\mathcal{T}_h$	0.005	0.01	0.02	0.03	0.04	0.05	0.1
800	0.029	0.066	0.128	0.153	0.178	0.219	0.319
3200	0.047	0.101	0.201	0.249	0.269	0.337	0.578
12800	0.106	0.194	0.277	0.412	0.502	0.594	0.663
51200	0.193	0.280	0.505	0.634	0.660	0.664	0.666
115200	0.243	0.419	0.635	0.664	0.666	0.666	0.666

TABLE 27 Example 12 ($d = 2$, Data 3) Values of Θ_{max} for indeterminate κ.

	$\theta_2 > 0$ $(\theta_1 = 0)$						
$\#\mathcal{T}_h$	0.005	0.01	0.02	0.03	0.04	0.05	0.1
384	0.111	0.119	0.153	0.153	0.153	0.153	0.435
1536	0.006	0.142	0.151	0.287	0.430	0.432	0.710
6144	0.138	0.229	0.423	0.497	0.571	0.642	0.854
24576	0.242	0.408	0.569	0.710	0.782	0.820	0.995
98304	0.393	0.587	0.765	0.854	0.924	0.962	0.999

TABLE 28 Example 12 ($d = 3$, Data 5) Values of Θ_{max} for indeterminate κ.

	$\theta_2 > 0$ $(\theta_1 = 0)$						
$\#\mathcal{T}_h$	0.005	0.01	0.02	0.03	0.04	0.05	0.1
336	0.009	0.019	0.029	0.039	0.058	0.068	0.147
2688	0.016	0.025	0.053	0.074	0.096	0.118	0.215
21504	0.025	0.049	0.099	0.154	0.192	0.206	0.378
172032	0.050	0.101	0.194	0.236	0.306	0.393	0.544

We conclude that if only μ contains indeterminacy, error indication is not affected much. However, if κ contains indeterminacy, error indication is affected. Especially on large meshes even small magnitudes of variation in κ may corrupt the process of error indication.

5 CONCLUSIONS

In the summary of this thesis we studied functional a posteriori error control in the context of the eddy-current problem.

In Section 3 we derived the classical functional minorants and majorants, and also derived an error equality for mixed approximations. All the presented estimates, and the equality, were tested numerically with several different examples in both two and three dimensions. We note that the error equality is computationally very cheap, since no auxiliary data needs to be computed. Furthermore, no constants are present either. Aside from giving the exact error we also showed with numerical examples that it also serves as a reliable error indicator which works nearly optimally.

In Section 4 we performed analysis on the effect of indeterminacy of the right hand side, and the material parameters. Computable quantities for practical simulations were presented and numerically tested. We emphasize that we were able to show that the indeterminacy of the right hand side can be exactly controlled. We also investigated by computational means how error indication is affected by indeterminacy of the material parameters. If only μ is indeterminate, error indication is not affected greatly. However, if κ is indeterminate, error indication is affected. Especially for large meshes even small amounts of variation in κ may corrupt the process of error indication.

Future research will concentrate on the error equality. As was already noted in [PIII, Section 2.1], the main application of the error equality may be in time-dependent static problems. Using, i.e., an implicit time-stepping scheme, one needs to solve at each time-step a problem where the error equality can be used.

YHTEENVETO (FINNISH SUMMARY)

Matemaattiset mallit tarjoavat korvaamattoman näkökulman luonnon ilmiöiden tulkitsemisessa, sekä myös teollisten tuotteiden ja prosessien optimoinnissa. Nämä matemaattiset mallit ovat usein osittaisdifferentiaaliyhtälöitä (ODY). Tällaisia yhtälöitä voidaan ratkaista tarkasti ainoastaan hyvin pelkistetyissä akateemisissa tapauksissa, joten realistisessa simulaatiossa ODY täytyy ratkaista laskennallisesti jollain numeerisella menetelmällä. Numeerinen ratkaisu on arvio tarkasta ratkaisusta, ja on olennaista tietää kuinka kaukana ne ovat toisistaan, eli kuinka paljon virhettä numeerinen ratkaisu sisältää. Numeerisen virheen arvioimiseen keskittyvää tieteenalaa kutsutaan a posteriori virhe-estimoinnin alaksi.

Tämä väitöskirja keskittyy tutkimaan Prof. Sergey Repinin kehittämiä funktionaalisia a posteriori virhe-estimaatteja. Väitöskirjaan liitetyissä artikkeleissa tutkitaan funktionaalisia estimaatteja pyörrevirtausyhtälölle, Stokesin yhtälölle, sekä diffuusioyhtälölle. Tämän väitöskirjan johdannossa käsitellään pyörrevirtausyhtälöä. Tämä yhtälö on toisen asteen ODY, joka on johdettu Maxwellin yhtälöistä. Maxwellin yhtälöillä ja pyörrevirtausyhtälöillä voidaan mallintaa elektromagneettisia ilmiöitä. Näillä yhtälöillä voidaan siis optimoida esimerkiksi antennin rakenne.

Luvussa 3 johdimme klassiset funktionaaliset ala- ja ylärajat joiden toimivuus todettiin myös numeerisilla testeillä. Tässä luvussa johdettiin myös uusi tulos, joka mahdollistaa tarkan virheen tuntemisen vaikka tarkkaa ratkaisua ei tunnettaisikaan. Näytimme että klassinen funktionaalinen yläraja antaa tarkan virheen ns. yhdistetyssä normissa. Varmistimme tuloksen toimivuuden myös numeerisilla testeillä, ja sitä kutsutaankin tässä työssä nimellä virheyhtäsuuruus. Sen lisäksi että klassinen yläraja antaa tarkan virheen, osoitimme myös että se toimii lähes optimaalisena virheindikaattorina yhdistetyssä normissa.

Luvussa 4 keskityimme tutkimaan sitä miten epävarma informaatio vaikuttaa numeeriseen simulointiin. Esimerkiksi jos simulaatiossa käytetty materiaalidata on saatu mittaamalla tätä materiaalia jollain työkalulla, se sisältää vähintäänkin tämän työkalun käytön takia syntyvän mittausvirheen verran virhettä. Epävarman datan vaikutus simulaatioihin on siten hyvin relevanttia. Käytännössä epävarma data generoi ratkaisujoukon, ja tutkimuksessa estimoidaan tämän ratkaisujoukon säteen suuruutta. Esitämme luvussa 4 kaksi käytännön simulaatioissa hyödyllistä työkalua, jotka myös testasimme numeerisesti. Tutkimme tässä luvussa myös epävarman materiaalidatan vaikutusta virheen indikointiin. Numeeristen testien perusteella vain toisella materiaaliparametrilla oli vaikutusta virheen indikointiin.

REFERENCES

[1] I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin, *Estimates of uncertainty errors for a magnetostatics problem*, D. Aubry, P. Díez, B. Tie, and N. Parés (Eds.), Proceedings and extended abstracts of ECCMAS ADMOS 2011, 65–69, 2011 (extended abstract).

[2] I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin, *A new error indicator for the Poisson problem*, R. Mäkinen, P. Neittaanmäki, T. Tuovinen, and K. Valpe (Eds.), Proceedings of 10th Finnish Mechanics Days, 324–330, 2009.

[3] I. Babuška and W. C. Rheinboldt, *A-posteriori error estimates for the finite element method*, Internat. J. Numer. Meth. Engrg., **12**: 1597–1615, 1978.

[4] I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal., **15**: 736–754, 1978.

[5] R. Beck, R. Hiptmair, R. Hoppe, and B. Wohlmuth, *Residual based a posteriori error estimators for eddy current computation*, Math. Model. Numer. Anal., **34**(1): 159–182, 2000.

[6] D. Braess, *Finite elements: theory, fast solvers, and applications in solid mechanics*, Cambridge University Press, New York, 1997.

[7] D. Braess and J. Schöberl, *Equilibrated residual error estimator for edge elements*, Math. Comp., **77**(262): 651–672, 2008.

[8] P. G. Ciarlet, *The finite element method for elliptic problems*, North Holland, New York, 1978.

[9] I. Ekeland and R. Temam, *Convex analysis and variational problems*, North Holland, Amsterdam, 1976.

[10] R. Eymard, T Gallouët, and R. Herbin, *Finite volume methods*, Handbook of Numerical Analysis, P. G. Ciarlet, J. L. Lions (Eds), **7**: 713–1020, 2000.

[11] N. Filonov, *On an inequality for the eigenvalues of the Dirichlet and Neumann problems for the Laplace operator*, St. Petersburg Math. J., **16**(2): 413–416, 2005.

[12] A. Hannukainen, *Functional type a posteriori error estimates for Maxwell's equations*, Proceedings of ENUMATH 2007 Conference, 41–48, 2008.

[13] I. Hlaváček, J. Chleboun, and I. Babuška, *Uncertain input data problems and the worst case scenario method*, Elsevier, Amsterdam, 2004.

[14] P. Houston, I. Perugia, and D. Schötzau, *An a posteriori error indicator for discontinuous Galerkin discretizations of $H(\mathrm{curl})$-elliptic partial differential equations*, J. Numer. Anal., **27**(1): 122–150, 2007.

[15] J. Jian-Ming, *The finite element method in electromagnetics*, Wiley, New York, 1993.

[16] F. Jochmann, *A compactness result for vector fields with divergence and curl in $L^q(\Omega)$ involving mixed boundary conditions*, Appl. Anal., **66**: 189–203, 1997.

[17] Y. Kuznetsov and S. Repin, *Guaranteed lower bounds of the smallest eigenvalues of elliptic differential operators*, J. Numer. Math., **21**(2): 135–156, 2013.

[18] O. A. Ladyzenskaja and N. N. Uraltseva, *Linear and Quasilinear Elliptic equations*, Academic Press, New York, 1968.

[19] O. Mali and S. Repin, *Estimates of accuracy limit for elliptic boundary value problems with uncertain data*, Adv. Math. Sci. Appl., **19**(2): 525–537, 2009.

[20] O. Mali and S. Repin, *Estimates of the indeterminacy set for elliptic boundary value problems with uncertain data*, J. Math. Sci., **150**(1): 1869–1874, 2008.

[21] O. Mali and S. Repin, *Two-sided estimates of the solution set for the reaction-diffusion problem with uncertain data*, Comput. Methods Appl. Sci., **15**: 183–198, 2010.

[22] O. Mali, S. Repin, and P. Neittaanmäki, *Accuracy verification methods: Theory and Algorithms*, Springer, 2014.

[23] The MathWorks, *Matlab*, <http://www.mathworks.com/products/matlab/> (1.11.2013).

[24] S. G. Mikhlin, *Constants in Some Inequalities of Analysis*, Wiley, 1986.

[25] P. Monk, *A posteriori error indicators for Maxwell's equations*, J. Comp. and Appl. Math., **100**: 173–190, 1998.

[26] P. Monk, *Finite element methods for Maxwell's equations*, Clarendon Press, Oxford, 2003.

[27] K. W. Morton and D. F. Mayers, *Numerical solution of partial differential equations: an introduction*, Cambridge University Press, New York, 2005.

[28] J. C. Nédélec, *A new family of mixed finite elements in $\mathbb{R}^3$*, Numeriche Matematik, **50**: 57–81, 1986.

[29] J. C. Nédélec, *Mixed finite elements in $\mathbb{R}^3$*, Numeriche Matematik, **35**: 315–341, 1980.

[30] P. Neittaanmäki and S. Repin, *Guaranteed error bounds for conforming approximations of a Maxwell type problem*, Computational Methods in Applied Sciences, **15**: 199–211, 2010.

[31] P. Neittaanmäki and S. Repin, *Reliable methods for computer simulation. Error control and a posteriori estimates*, Elsevier, Amsterdam, 2004.

[32] S. Nicaise, *On Zienkiewicz-Zhu error estimators for Maxwell's equations*, Comptes Rendus Mathematique, **340**(9): 697–702, 2005.

[33] D. Pauly, *On Maxwell's and Poincaré's constants*, Submitted, 2013.

[34] D. Pauly, *On Maxwell's constants in 3D*, Submitted, 2013.

[35] D. Pauly, *On the Maxwell inequalities for convex domains*, Submitted, 2013.

[36] D. Pauly and S. Repin, *Two-Sided A Posteriori Error Bounds for Electro-Magneto Static Problems*, Journal of Mathematical Sciences, **166**(1): 53–62, 2010.

[37] D. Pauly, S. Repin, and T. Rossi, *Estimates for deviations from exact solutions of the Cauchy problem for Maxwell's equations*, Annales Academiae Scientiarium Fennicae Mathematica, **36**: 661–676, 2011.

[38] L. E. Payne and H. F. Weinberger, *An optimal Poincaré inequality for convex domains*, Arch. Rational Mech. Anal., **5**: 286–292, 1960.

[39] T. Rahman and J. Valdman, *Fast MATLAB assembly of FEM matrices in 2D and 3D: nodal elements*, Applied mathematics and computation, **219**(13): 7151–7158, 2013.

[40] P. A. Raviart and J. M. Thomas, *Primal hybrid finite element methods for 2nd order elliptic equations*, Math. Comput. **31**: 391–413, 1977.

[41] S. Repin, *A posteriori estimates for partial differential equations*, Walter de Gruyter, Berlin, 2008.

[42] S. Repin, *Computable majorants of constants in the Poincaré and Friedrichs inequalities*, J. Math. Sci. **186**(2): 307–321, 2012.

[43] S. Repin, *Functional a posteriori estimates for Maxwell's equation*, J. Math. Sci., **142**(1): 1821–1827, 2007.

[44] S. Repin, S. Sauter, and A. Smolianski, *Two-sided a posteriori error estimates for mixed formulations of elliptic problems*, SIAM J. Numer. Anal., **45**(3): 928–945, 2007.

[45] G. I. Schuëller, *A state-of-the-art report on computational stochastic mechanics*, Prob. Engrg. Mech., **12**(4): 197–321, 1997.

[46] J. Valdman, *Minimization of Functional Majorant in A Posteriori Error Analysis Based on $H(\mathrm{div})$ Multigrid-Preconditioned CG Method*, Advances in Numerical Analysis, Article ID 164519, Volume 2009.

[47] O. C. Zienkiewicz and J. Z. Zhu, *A simple error estimator and adaptive procedure for practical engineering analysis*, Internat. J. Numer. Meth. Engrg., **24**: 337–357, 1987.

[48] O. C. Zienkiewicz and J. Z. Zhu, *Adaptive techniques in the finite element method*, Commun. Appl. Numer. Methods, **4**: 197–204, 1988.

ORIGINAL PAPERS

PI

A POSTERIORI ERROR ESTIMATES FOR A MAXWELL TYPE PROBLEM

by

I. Anjam, O. Mali, A. Muzalevsky, P. Neittaanmäki, and S. Repin (2009)

Russian Journal of Numerical Analysis and Mathematical Modelling,
24(5):395–408

Available online from De Gruyter Online:
http://www.degruyter.com/view/j/rnam.2009.24.issue-5/rjnamm.2009.025/rjnamm.2009.025.xml.

Russ. J. Numer. Anal. Math. Modelling, Vol. 24, No. 5, pp. 395–408 (2009)
DOI 10.1515/ RJNAMM.2009.025

A posteriori error estimates for a Maxwell type problem

I. ANJAM[*], O. MALI[*], A. MUZALEVSKY[†], P. NEITTAANMÄKI[*],
and S. REPIN[‡]

Abstract — In this paper, we discuss *a posteriori* estimates for the Maxwell type boundary-value problem. The estimates are derived by transformations of integral identities that define the generalized solution and are valid for any conforming approximation of the exact solution. It is proved analytically and confirmed numerically that the estimates indeed provide a computable and guaranteed bound of approximation errors. Also, it is shown that the estimates imply robust error indicators that represent the distribution of local (inter-element) errors measured in terms of different norms.

1. Introduction

In classical settings the Maxwell problem is defined by **E**, **D** (electric field and induction), **H** and **B** (magnetic field and induction) satisfying

$$\frac{\partial \mathbf{D}}{\partial t} - \operatorname{curl}\mathbf{H} = -\mathbf{J}$$

$$\frac{\partial \mathbf{B}}{\partial t} + \operatorname{curl}\mathbf{E} = 0$$

for all $(t,\mathbf{x})$ in $(0,T)\times\Omega$. Here Ω is a bounded and connected domain in $\mathbb{R}^d$ with the Lipschitz boundary $\partial\Omega$, and **J** is the applied current. Using the constituent relations

$$\mathbf{D} = \varepsilon\mathbf{E}$$
$$\mathbf{B} = \mu\mathbf{H}$$

where $\varepsilon(\mathbf{x}) > 0$ is the dielectric permittivity and $\mu(\mathbf{x}) > 0$ is the magnetic permeability (both μ and ε are assumed to be positive constants or positive bounded

[*] Department of Mathematical Information Technology, University of Jyvaskyla, Fi-40014, Finland

[†] Applied Mathematics Department, St. Petersburg State Polytechnic University, St. Petersburg 195251, Russia

[‡] St. Petersburg Department of V. A. Steklov Institute of Mathematics, St. Petersburg 191023, Russia

functions), we can rewrite the Maxwell equations in terms of **E** and **H** only:

$$\varepsilon \frac{\partial \mathbf{E}}{\partial t} - \operatorname{curl} \mathbf{H} = -\mathbf{J}$$
$$\mu \frac{\partial \mathbf{H}}{\partial t} + \operatorname{curl} \mathbf{E} = 0 .$$

These equations must be accompanied by initial conditions and suitable boundary conditions. In this paper, we assume that **E** satisfies the so-called PEC (perfect electric conductor) boundary condition

$$\mathbf{E} \times \mathbf{n} = 0 \qquad \text{on } \partial\Omega$$

where **n** denotes the unit outward normal to $\partial\Omega$. Usually the time derivatives are replaced by incremental relations. Using the backward-Euler scheme we have

$$\frac{\varepsilon}{\triangle t}\left(\mathbf{E}^n - \mathbf{E}^{n-1}\right) - \operatorname{curl} \mathbf{H}^n = -\mathbf{J}$$
$$\frac{\mu}{\triangle t}\left(\mathbf{H}^n - \mathbf{H}^{n-1}\right) + \operatorname{curl} \mathbf{E}^n = 0, \qquad n = 1, \dots, N, \quad N = \frac{T}{\triangle t}$$

where $\triangle t$ is the timestep. By eliminating $\mathbf{H}^n$ and transferring $\mathbf{E}^{n-1}$ and $\mathbf{H}^{n-1}$ to the right-hand side, we have

$$\operatorname{curl}\left(\mu^{-1} \operatorname{curl} \mathbf{E}^n\right) + \frac{\varepsilon}{(\triangle t)^2} \mathbf{E}^n = \frac{1}{\triangle t}\left(-\mathbf{J} + \frac{\varepsilon}{\triangle t} \mathbf{E}^{n-1} + \operatorname{curl} \mathbf{H}^{n-1}\right).$$

We denote the right-hand side by $\mathbf{f} \in L_2(\Omega, \mathbb{R}^d)$, set $\varkappa = \varepsilon(\triangle t)^{-2}$ and arrive at the model problem

$$\operatorname{curl}\left(\mu^{-1} \operatorname{curl} \mathbf{E}\right) + \varkappa \mathbf{E} = \mathbf{f} \qquad \text{in } \Omega \tag{1.1}$$
$$\mathbf{E} \times \mathbf{n} = 0 \qquad \text{on } \partial\Omega \tag{1.2}$$

in which the superscript n is omitted.

Below, we study (1.1)–(1.2) in the 2D case, so that the double curl is understood as $\underline{\operatorname{curl}} \operatorname{curl}$, where

$$\operatorname{curl} \mathbf{w} := \partial_1 w_2 - \partial_2 w_1 , \qquad \underline{\operatorname{curl}}\, \varphi := \begin{pmatrix} \partial_2 \varphi \\ -\partial_1 \varphi \end{pmatrix}.$$

We denote by $V(\Omega)$ the space $H(\operatorname{curl}; \Omega)$. This is a Hilbert space endowed with the norm

$$\|\mathbf{w}\|_{\operatorname{curl}} = \left(\|\mathbf{w}\|^2 + \|\operatorname{curl} \mathbf{w}\|^2\right)^{1/2}.$$

Here $||\cdot||$ is the L_2-norm of scalar- and vector-valued functions. By $V_0(\Omega)$ we denote a subspace of $V(\Omega)$ that consists of all the functions from V which satisfy boundary condition (1.2), i.e.,

$$V_0 := \{\mathbf{w} \in V \mid \mathbf{w} \times \mathbf{n} = 0 \text{ on } \partial\Omega\}.$$

The generalized solution $\mathbf{E} \in V_0$ of (1.1)–(1.2) is then defined by the integral relation

$$\int_\Omega \left(\mu^{-1} \operatorname{curl} \mathbf{E} \operatorname{curl} \mathbf{w} + \varkappa \mathbf{E} \cdot \mathbf{w}\right) \mathrm{d}\mathbf{x} = \int_\Omega \mathbf{f} \cdot \mathbf{w} \, \mathrm{d}\mathbf{x} \quad \forall \mathbf{w} \in V_0. \tag{1.3}$$

Also, we assume that $\mathbf{f}$ is a divergence-free function, so that

$$\int_\Omega \mathbf{f} \cdot \nabla\phi \, \mathrm{d}\mathbf{x} = 0 \quad \forall \phi \in \overset{\circ}{H}{}^1(\Omega)$$

and

$$0 < \mu_\ominus \leqslant \mu(\mathbf{x}) \leqslant \mu_\oplus .$$

Our goal is to obtain computable bounds of the difference between $\mathbf{E}$ and any function $\widetilde{\mathbf{E}} \in V_0$ measured in terms of the weighted (energy) norm

$$|[\mathbf{w}]|_{(\gamma,\delta)} := \int_\Omega \left(\gamma |\operatorname{curl} \mathbf{w}|^2 + \delta |\mathbf{w}|^2\right) \mathrm{d}\mathbf{x}.$$

A posteriori error estimation for the Maxwell's equations is a relatively new field of study. Most of the results that have been earlier obtained are based on the residual approach. In particular, residual type error estimates were studied in [1, 6, 7] and an equilibrated residual approach was presented in [2]. *A posteriori* estimates for non-conforming approximations for $H(\text{curl})$-elliptic partial differential equations were studied in [4]. A Zienkiewicz–Zhu type error estimate equivalent to the residual estimate in [1] was introduced in [5].

A posteriori estimates of the functional type present an efficient approach to the problem (a consequent exposition of the corresponding theory is given in [10, 11]). These estimates do not rely on any properties of the numerical method used to compute approximate solutions. This means that *a posteriori* estimates of the functional type are valid for any conforming approximation. Another important property of such estimates is that they do not contain mesh-dependent constants.

Functional type estimates for the Maxwell problem were derived in [3, 9, 12]. The equation (1.1) with $\varkappa > 0$ and $\varkappa = 0$ is considered in [12]. The upper bound for the case $\varkappa > 0$ does not contain a gap between the estimate and the true error (the estimate is sharp), but is sensitive with respect to small values of $\varkappa$. For the case $\varkappa = 0$ the sharpness of the presented upper bound cannot be proven. The same upper bound for the case $\varkappa > 0$ is presented in [3]. In addition, in [3] an upper bound for the case of complex $\varkappa$, $\mathscr{R}(\varkappa) \geqslant 0$ is presented. The sharpness of this upper bound cannot be proven. In [9] a sharp lower bound for $\varkappa > 0$ and two new upper bounds are presented. The first new upper bound is for $\varkappa \geqslant 0$ and it is insensitive with respect to small values of $\varkappa$. However, this estimate is sensitive with respect to large values of $\varkappa$ and the sharpness of this estimate cannot be proven. The second upper bound is derived in a more sophisticated way and provides a more general upper bound. Also, it behaves well with respect to small and large values of $\varkappa$.

In this paper, we derive functional *a posteriori* estimates for a model 2D problem that can be viewed as a simplified version of the Maxwell problem. As in [9], the derivation of estimates is based upon transformations of the corresponding integral identity. We prove that the estimates provide guaranteed and computable error bounds for the difference $\mathbf{E}-\widetilde{\mathbf{E}}$, where $\widetilde{\mathbf{E}}\in V_0$ is an approximation to the exact solution $\mathbf{E}$. In the last section, these theoretical results are confirmed by numerical experiments.

2. Error estimates

2.1. Upper bound of the error

First, we present some auxiliary results that are further used in the derivation of the upper bound.

By the Helmholtz decomposition of a vector-valued function, we represent the exact solution $\mathbf{E}$ as follows:

$$\mathbf{E}=\mathbf{E}_0+\nabla\psi$$

where $\mathbf{E}_0$ is a solenoidal vector-valued function and $\psi\in\overset{\circ}{H}{}^1(\Omega)$. Since $\operatorname{curl}\nabla\psi=0$, we rewrite (1.3) as follows:

$$\int_\Omega \mu^{-1}\operatorname{curl}\mathbf{E}_0\operatorname{curl}\mathbf{w}+\varkappa(\mathbf{E}_0+\nabla\psi)\cdot\mathbf{w}\,\mathrm{d}\mathbf{x}=\int_\Omega \mathbf{f}\cdot\mathbf{w}\ \mathrm{d}\mathbf{x}.$$

Next, we make the same decomposition for the trial function and set $\mathbf{w}=\mathbf{w}_0+\nabla\phi$. Since

$$\int_\Omega \mathbf{f}\cdot\nabla\phi\,\mathrm{d}\mathbf{x}=\int_\Omega \mathbf{E}_0\cdot\nabla\phi\,\mathrm{d}\mathbf{x}=\int_\Omega \mathbf{w}_0\cdot\nabla\psi\,\mathrm{d}\mathbf{x}=0$$

we observe that

$$\int_\Omega\left(\mu^{-1}\operatorname{curl}\mathbf{E}_0\operatorname{curl}\mathbf{w}_0+\varkappa\mathbf{E}_0\cdot\mathbf{w}_0+\varkappa\nabla\psi\cdot\nabla\phi\right)\mathrm{d}\mathbf{x}=\int_\Omega \mathbf{f}\cdot\mathbf{w}_0\,\mathrm{d}\mathbf{x}.$$

By setting $\mathbf{w}_0=0$ and $\phi=\psi$, we find that $\|\nabla\psi\|=0$. Hence, $\mathbf{E}$ is a divergence-free function.

Note that ϕ satisfies the relation

$$\int_\Omega \nabla\phi\cdot\nabla\xi\,\mathrm{d}\mathbf{x}=\int_\Omega \mathbf{w}\cdot\nabla\xi\,\mathrm{d}\mathbf{x}=-\int_\Omega(\operatorname{div}\mathbf{w})\xi\,\mathrm{d}\mathbf{x}\quad\forall\xi\in\overset{\circ}{H}{}^1(\Omega)$$

which implies the estimate

$$\|\nabla\phi\|\leqslant C_\Omega\|\operatorname{div}\mathbf{w}\| \tag{2.1}$$

where C_Ω is the constant in the Friedrichs inequality for the domain Ω. For solenoidal fields we also have the estimate (see, e.g., [7, 13])

$$\|\mathbf{w}_0\| \leqslant C_\Omega \|\mathrm{curl}\,\mathbf{w}_0\| = C_\Omega \|\mathrm{curl}\,\mathbf{w}\|\,. \tag{2.2}$$

Green's formula in a 2D setting states that for any $y \in H^1(\Omega)$ and any $\mathbf{w} \in H(\mathrm{curl};\Omega)$

$$\int_\Omega y\,\mathrm{curl}\,\mathbf{w}\,\mathrm{d}\mathbf{x} = \int_\Omega \underline{\mathrm{curl}}\,y \cdot \mathbf{w}\,\mathrm{d}\mathbf{x} + \int_{\partial\Omega} y(\mathbf{w}\times\mathbf{n})\,\mathrm{d}s$$

so we find that

$$\int_\Omega (\underline{\mathrm{curl}}\,y \cdot \mathbf{w} - y\,\mathrm{curl}\,\mathbf{w})\,\mathrm{d}\mathbf{x} = 0 \quad \forall \mathbf{w} \in V_0\,. \tag{2.3}$$

Proposition 2.1. *Let* $\widetilde{\mathbf{E}} \in V_0 \cap H(\mathrm{div};\Omega)$ *be an approximation of* $\mathbf{E}$*. For any* $y \in H^1(\Omega)$ *the following estimate holds:*

$$|[\mathbf{E}-\widetilde{\mathbf{E}}]|^2_{\gamma,\delta} \leqslant \mathscr{M}^2_\oplus(\lambda,\alpha_1,\alpha_2,\widetilde{\mathbf{E}},y) \tag{2.4}$$

with

$$\mathscr{M}^2_\oplus(\lambda,\alpha_1,\alpha_2,\widetilde{\mathbf{E}},y) := R_1(\lambda,\widetilde{\mathbf{E}},y) + \frac{\alpha_1}{4}R_2^2(\lambda,\widetilde{\mathbf{E}},y) + \frac{\alpha_2}{4}R_3^2(\lambda,\widetilde{\mathbf{E}},y)$$

where α_1 *and* α_2 *are arbitrary numbers in* $[1,+\infty)$ *and* $\varkappa$ *is a positive constant,*

$$\gamma = \left(1-\frac{1}{\alpha_1}\right)\mu^{-1}, \quad \delta = \left(1-\frac{1}{\alpha_2}\right)\varkappa$$
$$\lambda \in I_{[0,1]} := \{\lambda \in L^\infty(\Omega) \mid \lambda(\mathbf{x}) \in [0,1] \textit{ for a.e. } \mathbf{x}\in\Omega\}$$

and R_i*,* $i=1,2,3$*, are defined by* (2.9)–(2.11).

Proof. From (1.3) it follows that

$$\int_\Omega \left(\mu^{-1}\mathrm{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\,\mathrm{curl}\,\mathbf{w} + \varkappa(\mathbf{E}-\widetilde{\mathbf{E}})\cdot\mathbf{w}\right)\mathrm{d}\mathbf{x} = \int_\Omega \left(\mathbf{f}\cdot\mathbf{w} - \mu^{-1}\mathrm{curl}\,\widetilde{\mathbf{E}}\,\mathrm{curl}\,\mathbf{w} - \varkappa\widetilde{\mathbf{E}}\cdot\mathbf{w}\right)\mathrm{d}\mathbf{x}. \tag{2.5}$$

By (2.3) and (2.5) we obtain

$$\int_\Omega \left(\mu^{-1}\mathrm{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\,\mathrm{curl}\,\mathbf{w} + \varkappa(\mathbf{E}-\widetilde{\mathbf{E}})\cdot\mathbf{w}\right)\mathrm{d}\mathbf{x} = \int_\Omega \mathbf{r}(\widetilde{\mathbf{E}},y)\cdot\mathbf{w}\,\mathrm{d}\mathbf{x} + \int_\Omega d(\widetilde{\mathbf{E}},y)\,\mathrm{curl}\,\mathbf{w}\,\mathrm{d}\mathbf{x} \tag{2.6}$$

where

$$\begin{aligned}\mathbf{r}(\widetilde{\mathbf{E}},y) &:= \mathbf{f} - \underline{\mathrm{curl}}\, y - \varkappa\widetilde{\mathbf{E}}\\ d(\widetilde{\mathbf{E}},y) &:= y - \mu^{-1}\mathrm{curl}\,\widetilde{\mathbf{E}}\,.\end{aligned}$$

With the help of λ we decompose integral identity (2.6) as follows:

$$\begin{aligned}&\int_\Omega \Big(\mu^{-1}\mathrm{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\,\mathrm{curl}\,\mathbf{w} + \varkappa(\mathbf{E}-\widetilde{\mathbf{E}})\cdot\mathbf{w}\Big)\,\mathrm{d}\mathbf{x}\\ &\quad= \int_\Omega \lambda\,\mathbf{r}(\widetilde{\mathbf{E}},y)\cdot\mathbf{w}\,\mathrm{d}\mathbf{x} + \int_\Omega (1-\lambda)\mathbf{r}(\widetilde{\mathbf{E}},y)\cdot\mathbf{w}\,\mathrm{d}\mathbf{x} + \int_\Omega d(\widetilde{\mathbf{E}},y)\,\mathrm{curl}\,\mathbf{w}\,\mathrm{d}\mathbf{x} \qquad (2.7)\end{aligned}$$

where $\lambda \in I_{[0,1]}$. Since

$$\int_\Omega \lambda\,\mathbf{r}(\widetilde{\mathbf{E}},y)\cdot(\mathbf{E}-\widetilde{\mathbf{E}})\,\mathrm{d}\mathbf{x} \leqslant \left\|\frac{\lambda}{\varkappa^{1/2}}\mathbf{r}(\widetilde{\mathbf{E}},y)\right\| \|\varkappa^{1/2}(\mathbf{E}-\widetilde{\mathbf{E}})\|$$

and by inequalities (2.1) and (2.2)

$$\begin{aligned}&\int_\Omega (1-\lambda)\mathbf{r}(\widetilde{\mathbf{E}},y)\cdot(\mathbf{E}-\widetilde{\mathbf{E}})\,\mathrm{d}\mathbf{x}\\ &\qquad\qquad \leqslant \|(1-\lambda)\mathbf{r}(\widetilde{\mathbf{E}},y)\| \Big(C_\Omega\|\mathrm{div}\,\widetilde{\mathbf{E}}\| + C_\Omega\mu_\oplus^{1/2}\|\mu^{-1/2}\mathrm{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\|\Big)\,.\end{aligned}$$

By setting $\mathbf{w} = \mathbf{E}-\widetilde{\mathbf{E}}$ equation (2.7) becomes

$$\begin{aligned}&\int_\Omega \Big(\mu^{-1}|\mathrm{curl}(\mathbf{E}-\widetilde{\mathbf{E}})|^2 + \varkappa|\mathbf{E}-\widetilde{\mathbf{E}}|^2\Big)\,\mathrm{d}\mathbf{x}\\ &\qquad\qquad \leqslant R_1 + R_2\|\mu^{-1/2}\mathrm{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\| + R_3\|\varkappa^{1/2}(\mathbf{E}-\widetilde{\mathbf{E}})\| \qquad (2.8)\end{aligned}$$

where

$$\begin{aligned}R_1(\lambda,\widetilde{\mathbf{E}},y) &= C_\Omega\|(1-\lambda)\mathbf{r}(\widetilde{\mathbf{E}},y)\|\|\mathrm{div}\,\widetilde{\mathbf{E}}\| &(2.9)\\ R_2(\lambda,\widetilde{\mathbf{E}},y) &= C_\Omega\mu_\oplus^{1/2}\|(1-\lambda)\mathbf{r}(\widetilde{\mathbf{E}},y)\| + \|\mu^{1/2}d(\widetilde{\mathbf{E}},y)\| &(2.10)\\ R_3(\lambda,\widetilde{\mathbf{E}},y) &= \left\|\frac{\lambda}{\varkappa^{1/2}}\mathbf{r}(\widetilde{\mathbf{E}},y)\right\|\,. &(2.11)\end{aligned}$$

By applying Young's inequality to the right-hand side of (2.8), we obtain

$$\begin{aligned}&\int_\Omega \left(1-\frac{1}{\alpha_1}\right)\mu^{-1}|\mathrm{curl}(\mathbf{E}-\widetilde{\mathbf{E}})|^2\,\mathrm{d}\mathbf{x} + \int_\Omega \left(1-\frac{1}{\alpha_2}\right)\varkappa|\mathbf{E}-\widetilde{\mathbf{E}}|^2\,\mathrm{d}\mathbf{x}\\ &\qquad\qquad \leqslant R_1 + \frac{\alpha_1}{4}R_2^2 + \frac{\alpha_2}{4}R_3^2 \qquad (2.12)\end{aligned}$$

which implies (2.4).

Remark 2.1. A form of λ which is optimal (from the theoretical point of view) is obtained in [9], where similar estimates are considered for a 3D problem.

Corollary 2.1. If $\alpha_1 = \alpha_2 = 2$ then (2.4) comes in the form

$$|[\mathbf{E} - \widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)} \leqslant \mathscr{M}^{(\lambda)}_{\oplus} \tag{2.13}$$

where

$$\mathscr{M}^{(\lambda)}_{\oplus} := \mathscr{M}^2_{\oplus}(\lambda, \widetilde{\mathbf{E}}, y) = 2R_1(\lambda, \widetilde{\mathbf{E}}, y) + R^2_2(\lambda, \widetilde{\mathbf{E}}, y) + R^2_3(\lambda, \widetilde{\mathbf{E}}, y)$$

and this estimate is sharp.

Proof. It holds that

$$\inf_{\substack{\lambda \in I_{[0,1]} \\ y \in H^1(\Omega)}} \mathscr{M}^{(\lambda)}_{\oplus}(\widetilde{\mathbf{E}}, y) \leqslant \inf_{y \in H^1(\Omega)} \mathscr{M}^{(1)}_{\oplus}(\widetilde{\mathbf{E}}, y) \leqslant \mathscr{M}^{(1)}_{\oplus}(\widetilde{\mathbf{E}}, p)$$

where $p = \mu^{-1}\mathrm{curl}\,\mathbf{E}$. We have

$$\mathscr{M}^{(1)}_{\oplus}(\widetilde{\mathbf{E}}, p) = \|\mu^{-1/2}\mathrm{curl}(\mathbf{E} - \widetilde{\mathbf{E}})\|^2 + \|\varkappa^{1/2}(\mathbf{E} - \widetilde{\mathbf{E}})\|^2 = |[\mathbf{E} - \widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)}.$$

It means that the estimate is sharp.

Remark 2.2. By setting $\lambda = 1$ and $\lambda = 0$ we arrive at two particular forms of the error bound, which we call $\mathscr{M}^{(1)}_{\oplus}$ and $\mathscr{M}^{(0)}_{\oplus}$ respectively. They are as follows:

$$\mathscr{M}^{(1)}_{\oplus} = \|\varkappa^{-1/2}\mathbf{r}(\widetilde{\mathbf{E}}, y)\|^2 + \|\mu^{1/2}d(\widetilde{\mathbf{E}}, y)\|^2 \tag{2.14}$$

and

$$\mathscr{M}^{(0)}_{\oplus} = 2C_\Omega\|\mathbf{r}(\widetilde{\mathbf{E}}, y)\|\|\mathrm{div}\,\widetilde{\mathbf{E}}\| + \left(C_\Omega\mu^{1/2}_{\oplus}\|\mathbf{r}(\widetilde{\mathbf{E}}, y)\| + \|\mu^{1/2}d(\widetilde{\mathbf{E}}, y)\|\right)^2. \tag{2.15}$$

It should be noted that $\mathscr{M}^{(0)}_{\oplus}$ is well adapted to the case, in which $\varkappa$ is small (or even zero) and may lead to a considerable overestimation if $\varkappa$ is large. Conversely, $\mathscr{M}^{(1)}_{\oplus}$ is sensitive with respect to small $\varkappa$ and is well adapted to large values of this parameter. The combined majorant $\mathscr{M}^{(\lambda)}_{\oplus}$ is applicable to both cases. This property is due to the presence of the function λ, which allows us to compensate small values of $\varkappa$.

2.2. Lower bound of the error

Proposition 2.2. *Assume that* $\varkappa > 0$ *and* $\widetilde{\mathbf{E}} \in V_0$ *is an approximation of* $\mathbf{E}$*. For any* $\mathbf{w} \in V_0$ *the following estimate holds:*

$$|[\mathbf{E} - \widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)} \geqslant \mathscr{M}^2_{\ominus}(\widetilde{\mathbf{E}}, \mathbf{w}) \tag{2.16}$$

where

$$\mathscr{M}^2_{\ominus}(\widetilde{\mathbf{E}}, \mathbf{w}) := \int_\Omega \Big(2\mathbf{f}\cdot\mathbf{w} - \mu^{-1}|\operatorname{curl}\mathbf{w}|^2 \\ - \varkappa|\mathbf{w}|^2 - 2\mu^{-1}\operatorname{curl}\widetilde{\mathbf{E}}\operatorname{curl}\mathbf{w} - 2\varkappa\widetilde{\mathbf{E}}\cdot\mathbf{w}\Big)\mathrm{d}\mathbf{x}.$$

Proof. First, we note that

$$\begin{aligned}
&\sup_{w\in V_0}\int_\Omega\Big(\mu^{-1}\operatorname{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\operatorname{curl}\mathbf{w} + \varkappa\mathbf{w}\cdot(\mathbf{E}-\widetilde{\mathbf{E}}) \\
&\qquad - \frac{1}{2}(\mu^{-1}\operatorname{curl}\mathbf{w}\operatorname{curl}\mathbf{w} + \varkappa\mathbf{w}\cdot\mathbf{w})\Big)\mathrm{d}\mathbf{x} \\
&\leqslant \sup_{\substack{\tau\in H^1(\Omega,\mathbb{R})\\ \mathbf{w}\in L_2(\Omega,\mathbb{R}^2)}} \int_\Omega\Big(\mu^{-1}\operatorname{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\,\tau - \frac{1}{2}\mu^{-1}\tau\tau \\
&\qquad + \varkappa\mathbf{w}\cdot(\mathbf{E}-\widetilde{\mathbf{E}}) - \frac{1}{2}\varkappa\mathbf{w}\cdot\mathbf{w}\Big)\mathrm{d}\mathbf{x} = \frac{1}{2}|[\mathbf{E}-\widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&\sup_{w\in V_0}\int_\Omega\Big(\mu^{-1}\operatorname{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\operatorname{curl}\mathbf{w} + \varkappa\mathbf{w}\cdot(\mathbf{E}-\widetilde{\mathbf{E}}) \\
&\qquad - \frac{1}{2}(\mu^{-1}\operatorname{curl}\mathbf{w}\operatorname{curl}\mathbf{w} + \varkappa\mathbf{w}\cdot\mathbf{w})\Big)\mathrm{d}\mathbf{x} \\
&\geqslant \int_\Omega\Big(\mu^{-1}\operatorname{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\operatorname{curl}(\mathbf{E}-\widetilde{\mathbf{E}}) + \varkappa(\mathbf{E}-\widetilde{\mathbf{E}})\cdot(\mathbf{E}-\widetilde{\mathbf{E}}) \\
&\qquad - \frac{1}{2}\Big(\mu^{-1}|\operatorname{curl}(\mathbf{E}-\widetilde{\mathbf{E}})|^2 + \varkappa|\mathbf{E}-\widetilde{\mathbf{E}}|^2\Big)\Big)\mathrm{d}\mathbf{x} = \frac{1}{2}|[\mathbf{E}-\widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)}.
\end{aligned}$$

Thus, we conclude that

$$\begin{aligned}
\frac{1}{2}|[\mathbf{E}-\widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)} &= \sup_{\mathbf{w}\in V_0}\int_\Omega\Big(\mu^{-1}\operatorname{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\operatorname{curl}\mathbf{w} \\
&\qquad + \varkappa\mathbf{w}\cdot(\mathbf{E}-\widetilde{\mathbf{E}}) - \frac{1}{2}(\mu^{-1}\operatorname{curl}\mathbf{w}\operatorname{curl}\mathbf{w} + \varkappa\mathbf{w}\cdot\mathbf{w})\Big)\mathrm{d}\mathbf{x}.
\end{aligned}$$

Using equation (1.3), we obtain (2.16).

Corollary 2.2. The sharpest bound is given by the quantity

$$\mathscr{M}^2_{\ominus}(\widetilde{\mathbf{E}}) = \sup_{\mathbf{w}\in V_0} \mathscr{M}^2_{\ominus}(\widetilde{\mathbf{E}}, \mathbf{w}) .$$

By setting $\mathbf{w} = \mathbf{E} - \widetilde{\mathbf{E}}$, we find that

$$\mathscr{M}^2_{\ominus}(\widetilde{\mathbf{E}}) = |[\mathbf{E} - \widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)}$$

so the lower bound is sharp.

3. Numerical results

Estimates derived in the previous section have been verified in a series of numerical tests, which are discussed in this section. Approximations for the model problem were calculated with lowest-order Nédélec's elements of the first type (e.g., see [7, 8]). It should be noted that in the derivation of the upper bound we used the Helmholtz decomposition for the numerical approximation of the exact solution. Because of this, we are assuming that the numerical approximation belongs not only to $H(\mathrm{curl})$ but also to $H(\mathrm{div})$. With the lowest-order Nédélec's elements the normal component is not continuous across the element edges, so the divergence of approximate solutions is not square summable. To overcome this problem we chose to force the normal continuity by post-processing the numerical solution. Alternatively, one could use the nodal Courant elements to obtain approximate solutions, which belong to $H^1 \times H^1$. This problem does not arise with the upper bound $\mathscr{M}^{(1)}_{\oplus}$, because it can be derived separately without using Helmholtz decomposition (see [3, 9, 12]). Also the lower bound does not require the square summability of the divergence of the numerical approximation.

The free parameter y was obtained by globally minimizing the upper bounds with respect to y. Global minimization results in a finite element problem for y, which can be solved with standard nodal finite elements. Increasing the order of elements or using a more refined mesh than the mesh on which the approximate solution was computed results in better values for the upper bounds.

The performance of the upper bounds is measured by the so-called *efficiency index*

$$I_{\mathrm{eff}} = \left(\frac{\mathscr{M}^{(\lambda)}_{\oplus}}{|[\mathbf{E} - \widetilde{\mathbf{E}}]|^2_{(\mu^{-1},\varkappa)}} \right)^{1/2} .$$

To get sensible values for the lower bound, the free parameter $\mathbf{w}$ should be a better approximate solution to the problem than the original approximate solution $\mathbf{v}$. A better solution can be computed by simply refining the mesh and computing a new solution on this mesh. The finer the mesh, the better values for the lower bound we get.

Table 1.
Problem (3.3): Efficiency index values for different values of $\varkappa$.

	linear y			quadratic y		
$\varkappa$	$\mathscr{M}_{\oplus}^{(1)}$	$\mathscr{M}_{\oplus}^{(0)}$	$\mathscr{M}_{\oplus}^{(\lambda)}$	$\mathscr{M}_{\oplus}^{(1)}$	$\mathscr{M}_{\oplus}^{(0)}$	$\mathscr{M}_{\oplus}^{(\lambda)}$
10^{-3}	103.79	1.98	1.98	6.35	1.07	1.07
10^{-1}	10.42	1.98	1.98	1.18	1.07	1.06
10^{0}	3.42	1.98	1.91	1.02	1.08	1.02
10^{1}	1.42	1.96	1.42	1.00	1.18	1.00
10^{3}	1.00	7.14	1.00	1.00	7.05	1.00

Table 2.
Problem (3.3) with $\varkappa = 10^{-3}$: The sharpness of the upper bound $\mathscr{M}_{\oplus}^{(1)}$ and the lower bound $\mathscr{M}_{\ominus}$.

			linear y		quadratic y	
# elem	$\|[\mathbf{E}-\widetilde{\mathbf{E}}]\|^2$	$\mathscr{M}_{\ominus}^{2}$	$\mathscr{M}_{\oplus}^{(1)}$	I_{eff}	$\mathscr{M}_{\oplus}^{(1)}$	I_{eff}
82	0.11908		1897.90	126.25	7.04419	7.69
328	0.11908	0.08914	486.837	63.94	0.55972	2.17
1312	0.11908	0.11158	123.000	32.14	0.14689	1.11
5248	0.11908	0.11721	30.9403	16.12	0.12083	1.01

We are also interested in indicating the error distribution in different norms. The upper bound $\mathscr{M}_{\oplus}^{(1)}$ is the most suitable for this purpose, because it does not contain any constants. Using the two terms in $\mathscr{M}_{\oplus}^{(1)}$ separately we define the following error indicators

$$I_r(\widetilde{\mathbf{E}}, y) = \|\varkappa^{-1/2}\mathbf{r}(\widetilde{\mathbf{E}}, y)\| \tag{3.1}$$

$$I_d(\widetilde{\mathbf{E}}, y) = \|\mu^{1/2}d(\widetilde{\mathbf{E}}, y)\|\,. \tag{3.2}$$

By setting $y = \mu^{-1}\text{curl}\,\mathbf{E}$ we see that if the free parameter y is chosen properly, indicator (3.1) should give a good error distribution for the weighed L_2-norm of the error

$$\|\varkappa^{1/2}(\mathbf{E}-\widetilde{\mathbf{E}})\|\,.$$

Respectively, indicator (3.2) should give a good error distribution for the weighed $H(\text{curl})$-seminorm of the error

$$\|\mu^{-1/2}\text{curl}(\mathbf{E}-\widetilde{\mathbf{E}})\|\,.$$

For indicators (3.1) and (3.2) we also used a gradient averaging technique to compute y. It works as follows: for each node we calculate the approximate solution's curl values on the surrounding elements and weigh them by the sizes of respective elements. Then average the values to obtain a value for the node.

Table 3.
Problem (3.4): Efficiency index values with different mesh-sizes.

# elems	$\mathscr{M}_{\oplus}^{(1)}$	$\mathscr{M}_{\oplus}^{(0)}$	$\mathscr{M}_{\oplus}^{(\lambda)}$
72	1.00	1.05	1.00
246	1.00	1.04	1.00
980	1.00	1.02	1.00

For the *first test example* we take

$$\Omega = [0,1]^2\,, \quad \mu \equiv 1\,, \quad \varkappa > 0\,, \quad \mathbf{f} = (\pi^2 + \varkappa)\begin{pmatrix} \sin(\pi x_2) \\ \sin(\pi x_1) \end{pmatrix}. \tag{3.3}$$

For this problem we know the exact solution

$$\mathbf{u} = \begin{pmatrix} \sin(\pi x_2) \\ \sin(\pi x_1) \end{pmatrix}$$

which is the same for all $\varkappa > 0$. Table 1 shows the behaviour of the error majorants $\mathscr{M}_{\oplus}^{(1)}$, $\mathscr{M}_{\oplus}^{(0)}$, and $\mathscr{M}_{\oplus}^{(\lambda)}$ for different $\varkappa$. For each $\varkappa$ the approximate solution is calculated on a mesh with 82 elements and post-processed so that the divergence of the approximate solution becomes square summable. In the left-hand part of the table, the results correspond to the case in which y is computed by minimizing of majorants on the same mesh as for the approximate solution, using piecewise affine continuous approximation. The right-hand part exposes the results obtained by piece-wise quadratic approximations. It is not surprising that the efficiency indexes in the quadratic case are lower. The number of the degrees of freedom for quadratic approximation of y is roughly 4 times more than for the linear case. Another observation, which follows from Table 1 is that the majorants $\mathscr{M}_{\oplus}^{(1)}$ and $\mathscr{M}_{\oplus}^{(0)}$ may essentially overestimate the error, while $\mathscr{M}_{\oplus}^{(\lambda)}$ keeps small values of the efficiency index for all $\varkappa$. The dependence of upper bounds with respect to $\varkappa$ can also be seen in Fig. 1. The left picture corresponds to the linear approximation of y and the right picture corresponds to the quadratic approximation of y. From these results we also see that $\mathscr{M}_{\oplus}^{(1)}$ significantly benefits from using quadratic elements to approximate y.

Even though $\mathscr{M}_{\oplus}^{(1)}$ seriously overestimates the error with small values of $\varkappa$, the theory says that it is sharp. In principle, with $\mathscr{M}_{\oplus}^{(1)}$ one should be able to get as low efficiency index values as with $\mathscr{M}_{\oplus}^{(\lambda)}$. To verify this theory, we took the case $\varkappa = 10^{-3}$ and calculated the numerical approximation in a mesh with 82 elements. For this test we did not post-process the numerical approximation, because this majorant does not require that the approximate solution belongs to $H(\mathrm{div})$. To test the sharpness of this majorant, we calculated the free parameter y on subsequently refined meshes. The results in Table 2 agree with the theory. The convergence of the

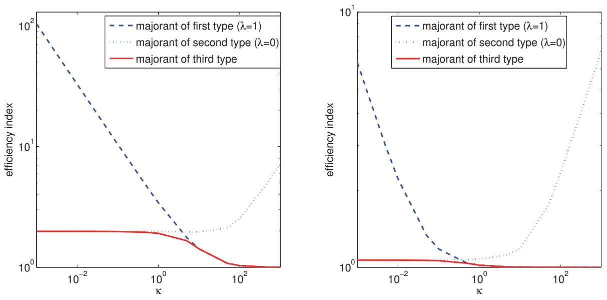


Figure 1. Problem (3.3): Efficiency indexes of the majorants $\mathscr{M}_{\oplus}^{(1)}$, $\mathscr{M}_{\oplus}^{(0)}$, and $\mathscr{M}_{\oplus}^{(\lambda)}$ for different $\varkappa$.

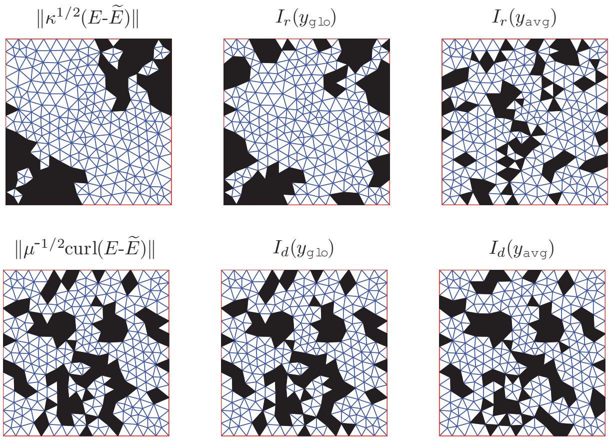

$\|\kappa^{1/2}(E\text{-}\widetilde{E})\|$
$I_r(y_{\mathtt{glo}})$
$I_r(y_{\mathtt{avg}})$
$\|\mu^{-1/2}\mathrm{curl}(E\text{-}\widetilde{E})\|$
$I_d(y_{\mathtt{glo}})$
$I_d(y_{\mathtt{avg}})$

Figure 2. Problem (3.3): Performance of error indicators.

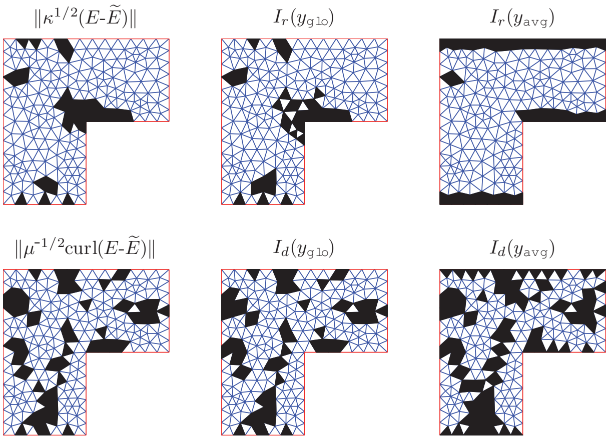

$\|\kappa^{1/2}(E\text{-}\widetilde{E})\|$
$I_r(y_{\mathtt{glo}})$
$I_r(y_{\mathtt{avg}})$
$\|\mu^{-1/2}\mathrm{curl}(E\text{-}\widetilde{E})\|$
$I_d(y_{\mathtt{glo}})$
$I_d(y_{\mathtt{avg}})$

Figure 3. Problem (3.4): Performance of error indicators.

linear y is slow, but using quadratic elements for y we clearly see that the upper bound converges to the exact error. Also, calculating the free parameter in the lower bound $\mathscr{M}_\ominus$ in the refined meshes shows that the lower bound is also sharp. From these results we can conclude that one can achieve arbitrary accuracy for the bounds if one is willing to use some time to compute the free parameters in the bounds.

For the *second test example* we take

$$\Omega = [0,1]^2 \setminus \left([\frac{1}{2},1] \times [0,\frac{1}{2}] \right), \quad \mu \equiv 1, \quad \varkappa = 1, \quad \mathbf{f} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \tag{3.4}$$

For this problem we do not know the exact solution. A reference solution was calculated in a mesh with 286114 elements. Table 3 gives the efficiency index values for the three upper bounds with some mesh sizes. An approximate solution was computed for each mesh and post-processed so that its divergence becomes square summable. The free parameter y was calculated with linear elements in the same mesh.

Figures 2 and 3 present the error indication results for indicators (3.1) and (3.2). Here, the function y was selected in two different ways: y_{glo} denotes the function obtained by global minimization of the majorant $\mathscr{M}_\oplus^{(1)}$, and y_{avg} denotes the function obtained by the simple averaging procedure described earlier. The free parameter y_{glo} was calculated with linear elements in the same mesh in which the approximate solution was calculated. In Figs. 1–3 we have marked with black color all elements with an error greater than the average error. The top rows present the results for indicator (3.1) and the bottom rows for indicator (3.2). In each row the first picture shows the exact error distribution that the indicators are supposed to indicate. The second picture shows the result of the indicator with y_{glo}, and the last picture shows the result for the same indicator with y_{avg}. Generally we observe good performance with y_{glo}. With y_{avg} the indicators do not perform so well.

References

1. R. Beck, R. Hiptmair, R. Hoppe and B. Wohlmuth, Residual based *a posteriori* error estimators for eddy current computation. *Math. Model. Numer. Anal.* (2000) **34**, No. 1, 159–182.
2. D. Braess and J. Schöberl, *Equilibrated residual error estimator for Maxwell's equation* (to appear).
3. A. Hannukainen, Functional type *a posteriori* error estimates for Maxwell's equations. *Proceedings of ENUMATH 2007 Conference*, 2008, pp. 41–48.
4. P. Houston, I. Perugia, and D. Schötzau, An *a posteriori* error indicator for discontinuous Galerkin discretizations of $H(\mathrm{curl})$-elliptic partial differential equations. *J. Numer. Anal.* (2007) **27**, No. 1, 122–150.
5. S. Nicaise, On Zienkiewicz–Zhu error estimators for Maxwell's equations. *Comptes Rendus Mathematique* (2005) **340**, No. 9, 697–702.
6. P. Monk, *A posteriori* error indicators for Maxwell's equations. *J. Comp. Appl. Math.* (1998) **100**, 173–190.
7. P. Monk, *Finite Element Methods for Maxwell's Equations*. Clarendon Press, Oxford, 2003.
8. J. C. Nédélec, Mixed Finite Elements in R^3. *Numeriche Matematik* (1980) **35**, 315–341.
9. P. Neittaanmäki and S. Repin, *Guaranteed error bounds for conforming approximations of a Maxwell type problem* (in print).
10. P. Neittaanmäki and S. Repin, *Reliable Methods for Computer Simulation. Error Control and a posteriori Estimates*. Elsevier, Amsterdam, 2004.
11. S. Repin, *A Posteriori* Estimates for Partial Differential Equations. *Walter de Gruyter, Berlin*, 2008.
12. S. Repin, Functional *a posteriori* estimates for Maxwell's equation. *J. Math. Sci. (N. Y.)* (2007) **142**, No. 1, 1821–1827.
13. J. Saranen, On an inequality of Friedrichs. *Math. Scand.* (1982) **51**, No. 2, 310 – 322.

PII

ON A POSTERIORI ERROR BOUNDS FOR APPROXIMATIONS OF THE GENERALIZED STOKES PROBLEM GENERATED BY THE UZAWA ALGORITHM

by

I. Anjam, M. Nokka, and S. Repin (2012)

Russian Journal of Numerical Analysis and Mathematical Modelling,
27(4):321–338

Available online from De Gruyter Online
http://www.degruyter.com/view/j/rnam.2012.27.issue-4/rnam-2012-0018/rnam-2012-0018.xml.

Russ. J. Numer. Anal. Math. Modelling, Vol. 27, No. 4, pp. 321–338 (2012)
DOI 10.1515/ rnam-2012-0018
© de Gruyter 2012

On *a posteriori* error bounds for approximations of the generalized Stokes problem generated by the Uzawa algorithm

I. ANJAM,[*] M. NOKKA,[†] and S. I. REPIN[‡]

Abstract — In this paper, we derive computable *a posteriori* error bounds for approximations computed by the Uzawa algorithm for the generalized Stokes problem. We show that for each Uzawa iteration both the velocity error and the pressure error are bounded from above by a constant multiplied by the L_2-norm of the divergence of the velocity. The derivation of the estimates essentially uses *a posteriori* estimates of the functional type for the Stokes problem.

1. Introduction

Let $\Omega \in \mathbb{R}^n$ be a bounded connected domain with a Lipschitz continuous boundary $\partial\Omega$. Henceforth, we use the space of vector valued functions

$$V(\Omega, \mathbb{R}^n) \;:=\; W_2^1(\Omega, \mathbb{R}^n)$$

and two spaces of tensor-valued functions

$$\begin{aligned} \Sigma(\Omega) \;&:=\; L_2(\Omega, \mathbb{M}^{n\times n}) \\ \Sigma(\mathrm{Div}, \Omega) \;&:=\; \{w \in \Sigma(\Omega) \mid \mathrm{Div}\, w \in L_2(\Omega, \mathbb{R}^n)\} \end{aligned}$$

where $\mathbb{M}^{n\times n}$ is the space of symmetric $n \times n$-matrices (tensors). The scalar product of tensors is denoted by two dots (:), and the L_2 norm of Σ is denoted by $\|\cdot\|_\Sigma$. The L_2 norm of scalar and vector valued functions is denoted by $\|\cdot\|$.

By $\mathring{S}(\Omega)$ we denote the closure of smooth solenoidal functions w with compact supports in Ω with respect to the norm $\|\nabla w\|_\Sigma$. Let $V_0(\Omega, \mathbb{R}^n)$ denote the subspace

[*] University of Jyväskylä, Department of Mathematical Information Technology, P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Finland. Email: immanuel.anjam@jyu.fi

[†] University of Jyväskylä, Department of Mathematical Information Technology, P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Finland. Email: marjaana.nokka@jyu.fi

[‡] V. A. Steklov Institute of Mathematics in St. Petersburg, Fontanka 27, RU-191024, St. Petersburg, Russia. Email: repin@pdmi.ras.ru

This work was supported by the Finnish foundation Emil Aaltosen säätiö and Russian Foundation for Basic Research Grants 11-01-00531a, 11-01-00324a.

of $V(\Omega,\mathbb{R}^n)$ that consists of functions with zero traces on $\partial\Omega$. The space of scalar valued square summable functions with zero mean is denoted by $\widetilde{L}_2(\Omega,\mathbb{R})$.

The classical statement of the generalized Stokes problem consists of finding a velocity field $u \in \mathring{S}(\Omega) + u_D$ and pressure $p \in \widetilde{L}_2(\Omega)$ which satisfy the relations

$$-\mathrm{Div}\,(\nu\nabla u) + \mu u + \nabla p = f \quad \text{in } \Omega \tag{1.1}$$

$$\mathrm{div}\, u = 0 \quad \text{in } \Omega \tag{1.2}$$

$$u = u_D \quad \text{on } \partial\Omega \tag{1.3}$$

where $f \in L_2(\Omega,\mathbb{R}^n)$, and

$$\int_{\partial\Omega} u_D \cdot n \,\mathrm{d}x = 0.$$

Here and later on n denotes the outward unit normal vector to $\partial\Omega$, and we assume that the material parameters ν and μ belong to the space $L_\infty(\Omega,\mathbb{R})$, and

$$0 < \underline{\nu} \leqslant \nu(x) \leqslant \overline{\nu}, \quad \forall x \in \overline{\Omega}$$
$$0 \leqslant \underline{\mu} \leqslant \mu(x) \leqslant \overline{\mu}, \quad \forall x \in \overline{\Omega}.$$

The generalized solution of (1.1)–(1.3) is a function $u \in \mathring{S}(\Omega) + u_D$ such that

$$\int_\Omega (\nu\nabla u : \nabla w + \mu u \cdot w)\,\mathrm{d}x = \int_\Omega f \cdot w \,\mathrm{d}x \quad \forall w \in \mathring{S}(\Omega). \tag{1.4}$$

It is well known that u can be defined as the first component of the saddle point problem generated by any of the Lagrangians

$$L(v,q) := \int_\Omega \left(\frac{1}{2}\nu|\nabla v|^2 + \frac{1}{2}\mu|v|^2 - q\,\mathrm{div}\, v - f\cdot v\right)\mathrm{d}x$$
$$L_A(v,q) := \int_\Omega \left(\frac{1}{2}\nu|\nabla v|^2 + \frac{1}{2}\mu|v|^2 + \frac{1}{2}\lambda|\mathrm{div}\, v|^2 - q\,\mathrm{div}\, v - f\cdot v\right)\mathrm{d}x.$$

The quantity in L_A is called the augmented Lagrangian (in which $\lambda \in \mathbb{R}_+$). We have

$$L(v,p) \leqslant L(u,p) \leqslant L(u,q) \quad \forall v \in V_0 + u_D,\ q \in L_2$$
$$L_A(v,p) \leqslant L_A(u,p) \leqslant L_A(u,q) \quad \forall v \in V_0 + u_D,\ q \in L_2.$$

From the right-hand side inequalities we see that $\int_\Omega (p-q)\mathrm{div}\, u\ \mathrm{d}x = 0$ for all $q \in L_2$, from which we conclude that $\mathrm{div}\, u = 0$. From the left-hand side inequalities it follows that for any solenoidal v we have $J(v) > J(u)$, where

$$J(v) := \int_\Omega \left(\frac{1}{2}\nu|\nabla v|^2 + \frac{1}{2}\mu|v|^2 - f\cdot v\right)\mathrm{d}x.$$

Indeed, the exact solution of the problems

$$\inf_{v\in V_0+u_D}\sup_{q\in L_2} L(v,q), \qquad \inf_{v\in V_0+u_D}\sup_{q\in L_2} L_A(v,q)$$

is (u,p). For a detailed exposition of this subject, we refer to [4].

Finding approximations of (u,p) can be performed by the Uzawa algorithm presented below.

Algorithm 1.1 (Uzawa algorithm).

```
1: Set k=0 and $\rho \in \mathbb{R}_+$. Make initial guess for $p^k \in \widetilde{L}_2$.
2: Find $u^k$ by minimizing the Lagrangian $L(v,p^k)$ or $L_A(v,p^k)$ w.r.t. $v$, i.e., by solving either (1.5) or (1.6), respectively.
   For the Lagrangian $L$, we have the problem: Find $u^k \in V_0 + u_D$ such that:
   $\int_\Omega \Big(\nu\nabla u^k : \nabla w + \mu u^k \cdot w\Big)\,\mathrm{d}x = \int_\Omega \Big(f\cdot w + p^k \operatorname{div} w\Big)\,\mathrm{d}x \quad \forall w \in V_0.$   (1.5)
   For the Lagrangian $L_A$, we have the problem: Find $u^k \in V_0 + u_D$ such that:
   $\int_\Omega \Big(\nu\nabla u^k : \nabla w + \mu u^k \cdot w + \lambda \operatorname{div} u^k \operatorname{div} w\Big)\,\mathrm{d}x$
       $= \int_\Omega \Big(f\cdot w + p^k \operatorname{div} w\Big)\,\mathrm{d}x \quad \forall w \in V_0.$   (1.6)
3: Find
   $p^{k+1} = (p^k - \rho \operatorname{div} u^k) \in \widetilde{L}_2.$   (1.7)
4: Set $k = k+1$ and go to step 2.
```

Our goal is to deduce computable bounds of the difference between u^k and the exact solution u in terms of the energy norms

$$|\!|\!| w |\!|\!|^2 := \int_\Omega \left(\nu|\nabla w|^2 + \mu|w|^2\right)\mathrm{d}x$$

and

$$]\!|\!| w |\!|\!|_\lambda^2 := \int_\Omega \left(\nu|\nabla w|^2 + \mu|w|^2 + \lambda|\operatorname{div} w|^2\right)\mathrm{d}x.$$

Theorem 1.1. *The Uzawa algorithm (Algorithm* 1.1*) converges, i.e.,*

$$u^k \overset{k\to\infty}{\longrightarrow} u \qquad \textit{strongly in } V(\Omega,\mathbb{R}^n)$$
$$p^k \overset{k\to\infty}{\longrightarrow} p \qquad \textit{weakly in } L_2(\Omega)$$

provided that

$$0 \;<\; \rho \;<\; 2\min(\underline{\nu},\underline{\mu}) \tag{1.8}$$

and $p^0 \in \widetilde{L}_2(\Omega)$*. If* $\mu \equiv 0$*, the condition is* $0 < \rho < 2\underline{\nu}$*. These conditions are the same for both* (1.5) *and* (1.6)*.*

Proof. The proof is based on well known arguments (see, e.g., [13]). However, for the convenience of the reader, we present the proof for the generalized Stokes problem, in the case of (1.5).

The exact solution of the generalized Stokes problem satisfies the relation

$$\int_\Omega (\nu\nabla u : \nabla w + \mu u\cdot w)\,\mathrm{d}x = \int_\Omega (f\cdot w + p\operatorname{div} w)\,\mathrm{d}x \quad \forall w \in V_0(\Omega). \tag{1.9}$$

We set $w = u^k - u$ and subtract (1.9) from (1.5), which gives

$$|\!|\!|\, u^k - u \,|\!|\!|^2 = \int_\Omega (p^k - p)\operatorname{div}(u^k - u)\mathrm{d}x.$$

Let $v^k := u^k - u$ and $q^k := p^k - p$. Then we rewrite this relation in the form

$$|\!|\!|\, v^k \,|\!|\!|^2 = \int_\Omega q^k \operatorname{div} v^k \,\mathrm{d}x. \tag{1.10}$$

On the other hand, (1.7) is equivalent to

$$\int_\Omega (p^{k+1} - p^k)\phi \,\mathrm{d}x + \rho\int_\Omega \operatorname{div} u^k \phi \,\mathrm{d}x = 0 \quad \forall \phi \in L_2(\Omega).$$

By setting $\phi = p^{k+1} - p$ we obtain

$$\int_\Omega (p^{k+1} - p^k)(p^{k+1} - p)\mathrm{d}x + \rho\int_\Omega \operatorname{div} u^k (p^{k+1} - p)\mathrm{d}x = 0$$

which is equivalent to

$$\int_\Omega (q^{k+1} - q^k)q^{k+1}\,\mathrm{d}x + \rho\int_\Omega \operatorname{div} v^k q^{k+1}\,\mathrm{d}x = 0$$

and

$$\|q^{k+1}\|^2 - \|q^k\|^2 + \|q^{k+1} - q^k\|^2 = -2\rho\int_\Omega \operatorname{div} v^k q^{k+1}\,\mathrm{d}x. \tag{1.11}$$

By combining (1.10) and (1.11), we obtain

$$\begin{aligned}
\|q^{k+1}\|^2 - \|q^k\|^2 + \|q^{k+1} - q^k\|^2 + 2\rho \,|\!|\!|\, v^k \,|\!|\!| \\
&= -2\rho\int_\Omega \operatorname{div} v^k (q^{k+1} - q^k)\mathrm{d}x \\
&\leqslant 2\rho\|\operatorname{div} v^k\|\,\|q^{k+1} - q^k\| \\
&\leqslant \delta^{-1}\rho^2\|\operatorname{div} v^k\|^2 + \delta\|q^{k+1} - q^k\|^2 \\
&\leqslant \delta^{-1}\rho^2\left(\|\nabla v^k\|_\Sigma^2 + \|v^k\|^2\right) + \delta\|q^{k+1} - q^k\|^2
\end{aligned} \tag{1.12}$$

where $\delta \in (0,1)$. Note that

$$|\!|\!| \, v^k \, |\!|\!|^2 \geqslant \underline{\nu}\|\nabla v^k\|_\Sigma^2 + \underline{\mu}\|v^k\|^2 \geqslant \min(\underline{\nu},\underline{\mu})\left(\|\nabla v^k\|_\Sigma^2 + \|v^k\|^2\right)$$

and, therefore, (1.12) implies the estimates

$$\|q^{k+1}\|^2 - \|q^k\|^2 + (1-\delta)\|q^{k+1} - q^k\|^2 \\ + \rho\left(2\min(\underline{\nu},\underline{\mu}) - \delta^{-1}\rho\right)\left(\|\nabla v^k\|_\Sigma^2 + \|v^k\|^2\right) \leqslant 0. \qquad (1.13)$$

Now, we sum inequalities (1.13) for $k = 0,\ldots,N$ and find that

$$\|q^{N+1}\|^2 + (1-\delta)\sum_{k=0}^{N}\|q^{k+1} - q^k\|^2 \\ + \rho\left(2\min(\underline{\nu},\underline{\mu}) - \delta^{-1}\rho\right)\sum_{k=0}^{N}\left(\|\nabla v^k\|_\Sigma^2 + \|v^k\|^2\right) \leqslant \|q^0\|. \qquad (1.14)$$

Because of condition (1.8), there exists a $\delta_* \in (0,1)$ such that

$$2\min(\underline{\nu},\underline{\mu}) - \delta_*^{-1}\rho > 0.$$

We set $\delta = \delta_*$ in (1.14), and see that

$$\|\nabla v^k\|_\Sigma^2 + \|v^k\|^2 = \|\nabla(u^k - u)\|_\Sigma^2 + \|u^k - u\|^2 \overset{k\to\infty}{\longrightarrow} 0.$$

Also, we see that $\|q^k\| = \|p^k - p\|$ is bounded in $L_2(\Omega)$, so $\|p^k\|$ is bounded in $L_2(\Omega)$. We also observe from (1.14), that

$$\|q^{k+1} - q^k\|^2 = \|p^{k+1} - p^k\|^2 \overset{k\to\infty}{\longrightarrow} 0$$

so we can extract from p^k a subsequence $p^{k'}$, which converges to some element p^* weakly in $L_2(\Omega)$. The equation (1.5) gives in the limit

$$\int_\Omega (\nu\nabla u : \nabla w + \mu u\cdot w)\,\mathrm{d}x = \int_\Omega (f\cdot w + p^*\operatorname{div} w)\,\mathrm{d}x \quad \forall w \in V_0$$

and by comparison to (1.9) we find that

$$\int_\Omega (p - p^*)\operatorname{div} w\,\mathrm{d}x = 0 \quad \forall w \in V_0$$

which means that $p^* = p + c$, where $c \in \mathbb{R}$. In other words, the sequence $p^{k'}$ converges weakly to p in $\widetilde{L}_2(\Omega)$ However, if $p^0 \in \widetilde{L}_2$, then it is easy to see from (1.7) that $p^k \in \widetilde{L}_2$ with all k. From this we make the conclusion that the sequence $p^{k'}$ converges weakly to p in $L_2(\Omega)$.

2. Error estimates for exact solutions generated by the Uzawa algorithm

In this section, we show that the errors of approximations generated by the Uzawa algorithm are controlled by the L_2-norm of the divergence of the velocity. First, we compare approximations computed on two consequent iterations and establish the following result.

Theorem 2.1. *Let (u^k, p^k) and (u^{k+1}, p^{k+1}) be the solutions of two consecutive iterations of the Uzawa algorithm. Then, for both* (1.5) *and* (1.6) *we have*

$$\begin{aligned} |\!|\!|\, u^{k+1} - u^k \,|\!|\!| \;&\leqslant\; \sqrt{\underline{\nu}}^{-1}\rho\,\|\operatorname{div} u^k\| && (2.1)\\ \|p^{k+1} - p^k\| \;&=\; \rho\,\|\operatorname{div} u^k\|. && (2.2) \end{aligned}$$

In addition, for (1.6) *we also have*

$$]\!|\!|\, u^{k+1} - u^k \,|\!|\![_\lambda \leqslant \sqrt{\underline{\nu}}^{-1}\rho\,\|\operatorname{div} u^k\|. \tag{2.3}$$

Proof. The equation for pressure (2.2) follows directly from (1.7). By subtracting the kth equation (1.5) from the $(k+1)$th equation, we obtain

$$\int_\Omega \nu\nabla(u^{k+1} - u^k) : \nabla w + \mu(u^{k+1} - u^k)\cdot w \,\mathrm{d}x = \int_\Omega (p^{k+1} - p^k)\operatorname{div} w \,\mathrm{d}x.$$

Since

$$\|\operatorname{div} w\| \leqslant \|\nabla w\|_\Sigma \leqslant \sqrt{\underline{\nu}}^{-1}\|\sqrt{\nu}\nabla w\|_\Sigma \leqslant \sqrt{\underline{\nu}}^{-1}\, |\!|\!|\, w \,|\!|\!|$$

we can estimate the right-hand side with

$$\begin{aligned} \int_\Omega (p^{k+1} - p^k)\operatorname{div} w \,\mathrm{d}x \;&\leqslant \|p^{k+1} - p^k\|\,\|\operatorname{div} w\| \\ &\leqslant \sqrt{\underline{\nu}}^{-1}\|p^{k+1} - p^k\|\; |\!|\!|\, w \,|\!|\!| \,. \end{aligned}$$

By choosing $w = u^{k+1} - u^k$, we obtain

$$|\!|\!|\, u^{k+1} - u^k \,|\!|\!|^2 \leqslant \sqrt{\underline{\nu}}^{-1}\|p^{k+1} - p^k\|\; |\!|\!|\, u^{k+1} - u^k \,|\!|\!| \,.$$

By (2.2) we obtain the estimate for velocity (2.1). The estimate (2.3) is obtained with exactly the same arguments applied for the augmented form (1.6). Since $|\!|\!|\, w \,|\!|\!| \leqslant\,]\!|\!|\, w \,|\!|\![_\lambda$ for all $\lambda \in \mathbb{R}_+$, we see by (2.3), that the estimate (2.1) holds also for approximations calculated by (1.6).

Henceforth, we will use functional *a posteriori* error estimates for the Stokes problem derived in [11, 12]. For a consequent exposition of the theory of functional *a posteriori* error estimates we refer the reader to [8, 10].

The following lemma is essential in deriving our main results.

Lemma 2.1. *Let Ω be a bounded domain with Lipschitz continuous boundary $\partial\Omega$. Then there exists a positive constant C_{LBB} depending on the domain Ω such that for any function $g \in \widetilde{L}_2(\Omega)$ there is a function $v \in V_0$ satisfying the condition* $\operatorname{div} v = g$*, and*

$$\|\nabla v\|_\Sigma \leqslant C_{\text{LBB}}^{-1}\|g\|.$$

Here C_{LBB} is the constant in the well-known Ladyzhenskaya-Babuška-Brezzi (LBB) condition (*see, e.g.,* [1,2]). *See proof in* [6,7].

For some simple domains the constant C_{LBB}, or the bounds for it, are known (see, e.g., [3,5,9]).

Lemma 2.1 implies an important corollary. Let $v \in V_0$, and $\operatorname{div} v = g$. Then there exists a function $v_g \in V_0$ such that $\operatorname{div}(v - v_g) = 0$, and

$$\|\nabla v_g\|_\Sigma \leqslant C_{\text{LBB}}^{-1}\|g\| = C_{\text{LBB}}^{-1}\|\operatorname{div} v\|.$$

This means that there exists a solenoidal field $v_0 = (v - v_g) \in \overset{\circ}{S}(\Omega)$ such that

$$\|\nabla(v - v_0)\|_\Sigma \leqslant C_{\text{LBB}}^{-1}\|\operatorname{div} v\|.$$

A similar estimate holds for $v \in V_0 + u_D$. Indeed, for $v - u_D$ we can find a solenoidal field $v_0 \in \overset{\circ}{S}(\Omega)$ such that

$$\|\nabla(v - u_D - v_0)\|_\Sigma \leqslant C_{\text{LBB}}^{-1}\|\operatorname{div}(v - u_D)\| \leqslant C_{\text{LBB}}^{-1}\|\operatorname{div} v\|.$$

Thus, we can find a function $w_0 \in \overset{\circ}{S}(\Omega) + u_D$ such that

$$\|\nabla(v - w_0)\|_\Sigma \leqslant C_{\text{LBB}}^{-1}\|\operatorname{div} v\|. \tag{2.4}$$

With the help of (2.4) we can now derive our main results. We show that the errors of u^k and p^k generated on the iteration k of the Uzawa algorithm are both estimated from above by the L_2-norm of the divergence of u^k multiplied by a constant depending on C_{LBB}. The proofs are based on the derivation of functional *a posteriori* error estimates for the generalized Stokes problem as they are presented in [12].

Theorem 2.2. *Let u^k be the exact solution computed on the iteration k of the Uzawa algorithm. Then, for solutions calculated by* (1.5) *or* (1.6)*, we have*

$$|\!|\!| u - u^k |\!|\!| \leqslant 2C\|\operatorname{div} u^k\| \tag{2.5}$$

where

$$C := C_{\text{LBB}}^{-1}\sqrt{C_{\text{F}}^2\overline{\mu} + \overline{\nu}}. \tag{2.6}$$

Here C_{F} is the constant in the Friedrichs inequality

$$\|w\| \leqslant C_{\text{F}}\|\nabla w\|_\Sigma$$

and C_{LBB} is the constant in the LBB*-condition.*

Proof. Let $u_0 \in \mathring{S}(\Omega) + u_D$ be such that, by using (2.4), we have

$$\|\nabla(u^k - u_0)\|_\Sigma \leqslant C_{\mathrm{LBB}}^{-1}\|\mathrm{div}\, u^k\|. \tag{2.7}$$

Let the pair (u^k, p^k) be an approximation of the saddle point computed on the iteration k. We can now write

$$|\!|\!|\, u - u^k \,|\!|\!| \leqslant |\!|\!|\, u - u_0 \,|\!|\!| + |\!|\!|\, u_0 - u^k \,|\!|\!| \,. \tag{2.8}$$

First, we estimate from above the first term on the right-hand side of (2.8). Let $w \in \mathring{S}$. By subtracting the integral $\int_\Omega (\nu\nabla u_0 : \nabla w + \mu u_0 \cdot w)\,\mathrm{d}x$ from both sides of (1.4) we obtain

$$\int_\Omega (\nu\nabla(u - u_0) : \nabla w + \mu(u - u_0)\cdot w)\,\mathrm{d}x \\ = \int_\Omega ((f - \mu u_0)\cdot w - \nu\nabla u_0 : \nabla w)\,\mathrm{d}x. \tag{2.9}$$

It is easy to see that

$$\int_\Omega (\mathrm{Div}\,\tau\cdot w + \tau : \nabla w)\,\mathrm{d}x = 0 \quad \forall \tau \in \Sigma(\mathrm{Div}, \Omega),\;\; w \in V_0(\Omega) \tag{2.10}$$

and

$$\int_\Omega (\nabla q\cdot w + q\,\mathrm{div}\, w)\,\mathrm{d}x = 0 \quad \forall q \in W_2^1(\Omega, \mathbb{R}),\;\; w \in V_0(\Omega).. \tag{2.11}$$

By adding (2.10) and (2.11) to the right-hand side of (2.9), we rewrite it in the form

$$\int_\Omega ((f - \mu u_0 + \mathrm{Div}\,\tau - \nabla q)\cdot w + (\tau - \nu\nabla u_0) : \nabla w)\,\mathrm{d}x \tag{2.12}$$

which is equivalent to

$$\int_\Omega \Big(\Big(f - \mu u^k + \mathrm{Div}\,\tau - \nabla q\Big)\cdot w + \Big(\tau - \nu\nabla u^k\Big) : \nabla w\Big)\,\mathrm{d}x \\ + \int_\Omega \Big(\nu\nabla(u^k - u_0) : \nabla w + \mu(u^k - u_0)\cdot w\Big)\,\mathrm{d}x. \tag{2.13}$$

Let us choose $\tau = \nu\nabla u^k$ and $q = p^k$. In view of (1.5), we see that that the first integral of (2.13) vanishes. Indeed,

$$\int_\Omega \Big(\Big(f - \mu u^k + \mathrm{Div}\,\nu\nabla u^k - \nabla p^k\Big)\cdot w \;+ \Big(\nu\nabla u^k - \nu\nabla u^k\Big) : \nabla w\Big)\,\mathrm{d}x \\ = \int_\Omega \Big(f\cdot w + p^k\,\mathrm{div}\, w - \nu\nabla u^k : \nabla w - \mu u^k\cdot w\Big)\,\mathrm{d}x = 0. \tag{2.14}$$

Since w is a function from $\mathring{S}$, the same conclusion is also true if u^k has been calculated by (1.6). We combine (2.9) with (2.12)–(2.14), and arrive at the relation

$$\begin{aligned}&\int_\Omega (\nu\nabla(u-u_0):\nabla w+\nu(u-u_0)\cdot w)\,\mathrm{d}x\\ &\qquad\qquad=\int_\Omega \Big(\nu\nabla(u^k-u_0):\nabla w+\mu(u^k-u_0)\cdot w\Big)\,\mathrm{d}x.\end{aligned} \tag{2.15}$$

The right-hand side of (2.15) can be estimated from above as follows:

$$\begin{aligned}&\int_\Omega \big(\nu\nabla(u^k-u_0):\nabla w+\mu(u^k-u_0)\cdot w\big)\mathrm{d}x\\ &\qquad=\int_\Omega \Big(\sqrt{\nu}\nabla(u^k-u_0):\sqrt{\nu}\nabla w+\sqrt{\mu}(u^k-u_0)\cdot\sqrt{\mu}w\Big)\,\mathrm{d}x\\ &\qquad\leqslant \|\sqrt{\nu}\nabla(u^k-u_0)\|_\Sigma\|\sqrt{\nu}\nabla w\|_\Sigma+\|\sqrt{\mu}(u^k-u_0)\|\,\|\sqrt{\mu}w\|\\ &\qquad\leqslant |\!|\!|\, u^k-u_0 \,|\!|\!|\;|\!|\!|\, w \,|\!|\!|\end{aligned} \tag{2.16}$$

where we have used the Cauchy–Schwarz inequality. We set $w=u-u_0$, and find that

$$|\!|\!|\, u-u_0 \,|\!|\!| \leqslant |\!|\!|\, u^k-u_0 \,|\!|\!|\;. \tag{2.17}$$

Note that for all $w\in V$ we have

$$\begin{aligned}|\!|\!|\, w \,|\!|\!|^2 \;&=\; \|\sqrt{\nu}\nabla w\|_\Sigma^2+\|\sqrt{\mu}w\|^2\\ &\leqslant\; \overline{\nu}\|\nabla w\|_\Sigma^2+\overline{\mu}\|w\|^2\\ &\leqslant\; \overline{\nu}\|\nabla w\|_\Sigma^2+C_{\mathrm{F}}^2\overline{\mu}\|\nabla w\|_\Sigma^2\\ &\leqslant\; \big(C_{\mathrm{F}}^2\overline{\mu}+\overline{\nu}\big)\,\|\nabla w\|_\Sigma^2.\end{aligned} \tag{2.18}$$

We substitute (2.17) into (2.8), and use (2.18) with $w=u-u_0$, and obtain

$$\begin{aligned}|\!|\!|\, u-u^k \,|\!|\!| \;&\leqslant\; 2\,|\!|\!|\, u_0-u^k \,|\!|\!|\\ &\leqslant\; 2\sqrt{C_{\mathrm{F}}^2\overline{\mu}+\overline{\nu}}\|\nabla(u_0-u^k)\|_\Sigma.\end{aligned} \tag{2.19}$$

Now, (2.7) and (2.19) imply the estimate

$$|\!|\!|\, u-u^k \,|\!|\!| \leqslant 2C_{\mathrm{LBB}}^{-1}\sqrt{C_{\mathrm{F}}^2\overline{\mu}+\overline{\nu}}\|\mathrm{div}\,u^k\|=2C\|\mathrm{div}\,u^k\|$$

where C is defined in (2.6).

In order to prove a similar estimate for the pressure, we also need Lemma 2.1. Let $q\in\widetilde{L}_2$ be an approximation of the exact pressure p. Then $(p-q)\in\widetilde{L}_2$ and there exists a function $\overline{w}\in V_0$ such that

$$\mathrm{div}(\overline{w})=p-q \tag{2.20}$$

and

$$\|\nabla\overline{w}\|_\Sigma \leqslant C_{\mathrm{LBB}}^{-1}\|p-q\|. \tag{2.21}$$

Theorem 2.3. *Let p^k be the function computed on the iteration k of the Uzawa algorithm. Then,*

$$\|p-p^k\| \;\leqslant\; \mathbb{C}\|\operatorname{div} u^k\| \tag{2.22}$$

where $\mathbb{C} = 2C^2$ for (1.5)*, and $\mathbb{C} = 2C^2+\lambda$ for* (1.6).

Proof. We use (2.20) for $q = p^k$ and obtain

$$\|p-p^k\|^2 = \int_\Omega \operatorname{div}\overline{w}(p-p^k)\mathrm{d}x = \int_\Omega \operatorname{div}\overline{w}\,p + \nabla p^k\cdot\overline{w}\,\mathrm{d}x.$$

Multiplying (1.1) by $\overline{w}$ and integrating over Ω, we obtain

$$\int_\Omega \operatorname{div}\overline{w}\,p\;\mathrm{d}x \;=\int_\Omega (\nu\nabla u:\nabla\overline{w}+\mu u\cdot\overline{w}-f\cdot\overline{w})\,\mathrm{d}x.$$

In view of this relation, we have

$$\|p-p^k\|^2 = \int_\Omega \Big(\nu\nabla u:\nabla\overline{w}+\mu u\cdot\overline{w}-f\cdot\overline{w}+\nabla p^k\cdot\overline{w}\Big)\,\mathrm{d}x.$$

We use (2.10) with $w=\overline{w}$, and arrive at the relation

$$\begin{aligned}\|p-p^k\|^2 = &\int_\Omega \Big(\Big(-f+\mu u^k-\operatorname{Div}\tau+\nabla p^k\Big)\cdot\overline{w}+\Big(\nu\nabla u^k-\tau\Big):\nabla\overline{w}\Big)\,\mathrm{d}x\\ &+\int_\Omega\Big(\nu\nabla(u-u^k):\nabla\overline{w}+\mu(u-u^k)\cdot\overline{w}\Big)\,\mathrm{d}x.\end{aligned} \tag{2.23}$$

As before, we choose $\tau=\nu\nabla u^k$, and observe that the first integral is zero. By estimating the latter integral with the help of the same arguments as in (2.16), we find that

$$\|p-p^k\|^2 \;\leqslant |\!|\!|\; u-u^k \;|\!|\!|\;|\!|\!|\;\overline{w}\;|\!|\!|\;. \tag{2.24}$$

By (2.18) and (2.21), we obtain

$$\begin{aligned}|\!|\!|\;\overline{w}\;|\!|\!|^2 \;&\leqslant\; \big(C_{\mathrm{F}}^2\overline{\mu}+\overline{\nu}\big)\,\|\nabla\overline{w}\|_\Sigma^2\\ &\leqslant\; C_{\mathrm{LBB}}^{-2}\big(C_{\mathrm{F}}^2\overline{\mu}+\overline{\nu}\big)\,\|p-p^k\|^2\\ &=\; C^2\|p-p^k\|^2\end{aligned} \tag{2.25}$$

where C is defined in (2.6). Substituting (2.25) into (2.24) results in the estimate

$$\|p-p^k\| \;\leqslant\; C\;|\!|\!|\; u-u^k\;|\!|\!|\;.$$

Now, we apply Theorem 2.2 and deduce (2.22).

In the case of (1.6), we add

$$\int_\Omega \lambda \operatorname{div}(u^k - u^k)\operatorname{div}\overline{w}\,\mathrm{d}x = 0$$

to (2.23) and obtain

$$\begin{aligned}\|p-p^k\|^2 = &\int_\Omega \Big(\Big(-f+\mu u^k - \operatorname{Div}\tau + \nabla p^k\Big)\cdot\overline{w} + \lambda \operatorname{div} u^k \operatorname{div}\overline{w}\Big)\,\mathrm{d}x\\ &+\int_\Omega \Big(\nu\nabla u^k - \tau\Big) : \nabla\overline{w}\,\mathrm{d}x\\ &+\int_\Omega \Big(\nu\nabla(u-u^k):\nabla\overline{w} + \mu(u-u^k)\cdot\overline{w} - \lambda\operatorname{div}u^k\operatorname{div}\overline{w}\Big)\,\mathrm{d}x.\end{aligned}$$

Again, we choose $\tau = \nu\nabla u^k$, and see from (1.6) that the first and second integrals are zero. By estimating the latter integral with same arguments as in (2.16), we obtain

$$\|p-p^k\|^2 \leqslant |\!|\!| u-u^k |\!|\!|\, |\!|\!| \overline{w} |\!|\!| + \lambda\|\operatorname{div}u^k\|\,\|\operatorname{div}\overline{w}\|. \tag{2.26}$$

Recall that $\operatorname{div}\overline{w} = p - p^k$. Now, (2.25) and (2.26) imply the estimate

$$\|p-p^k\| \leqslant C\,|\!|\!| u-u^k |\!|\!| + \lambda\|\operatorname{div}u^k\|.$$

Applying Theorem 2.2 results in (2.22).

By Theorems 2.2 and 2.3, we easily conclude the following statement.

Remark 2.1. The classical Stokes problem corresponds to the case where $\mu \equiv 0$ and ν is a constant. Let (u^k, p^k) be the exact solution computed on the iteration k of the Uzawa algorithm, for the Stokes problem. Then, for velocity we have (for both cases (1.5) and (1.6))

$$\|\nabla(u-u^k)\| \leqslant 2C_{\mathrm{LBB}}^{-1}\|\operatorname{div}u^k\|.$$

For the pressure we have

$$\|p-p^k\| \leqslant \tilde{\mathbb{C}}\|\operatorname{div}u^k\|$$

where $\tilde{\mathbb{C}} = 2C_{\mathrm{LBB}}^{-2}\nu$ for (1.5) and $\tilde{\mathbb{C}} = 2C_{\mathrm{LBB}}^{-2}\nu + \lambda$ for (1.6).

3. Computable error estimates for approximations generated by the Uzawa algorithm

Let $\mathscr{T}_h$ be a mesh having the characteristic size h, and let the spaces $V_{0h}(\Omega,\mathbb{R}^n)$ and $Q_h(\Omega)$ be finite dimensional subspaces of $V_0(\Omega,\mathbb{R}^n)$ and $\widetilde{L}_2(\Omega)$, respectively. We assume that for all $v_h \in V_{0h} + u_D$ it holds that $\operatorname{div} v_h \in Q_h$. We also assume that the spaces are constructed so that they satisfy the discrete LBB-condition, i.e, for any $q_h \in Q_h$ with zero mean, there exists $v_h \in V_{0h}$ such that

$$\operatorname{div} v_h = q_h$$

and

$$\|\nabla v_h\|_\Sigma \leqslant c\|q_h\|$$

where the positive constant c does not depend on h.

Let $u_h^k \in V_{0h} + u_D$ be an approximation of u^k calculated on the mesh $\mathscr{T}_h$. We need to combine the error of the pure Uzawa algorithm with the approximation error. Below we present the corresponding results, where we set $p^k = p_h^k \in Q_h$ on the iteration k, and understand u^k as satisfying (1.5), or (1.6), with the chosen p_h^k. Then, the pair (u^k, p_h^k) can be viewed as the exact pair associated with the Uzawa algorithm on iteration k.

Our first goal is to derive fully computable error majorants $M_\oplus^k$ and $M_\oplus^{k,\lambda}$ for approximate solutions (e.g., u_h^k) of the problems generated at the first step of Uzawa algorithm by the Lagrangians L and L_A, respectively. In order to make the quality of the majorants robust with respect to small or large values of the material functions ν or μ, we apply the same method that was suggested in [12] for the generalized Stokes problem.

Later we combine these estimates with the estimates of the difference between u and u^k and obtain estimates applicable for approximate solutions computed within the framework of finite dimensional approximations.

First, we prove the following result for the problem generated by the Lagrangian L.

Theorem 3.1. *Let* (u^k, p_h^k) *be the exact solution on the iteration k of the Uzawa algorithm. Then, for the solutions calculated by* (1.5)*, and for an approximation* $u_h^k \in V_{0h} + u_D$ *we have*

$$\|\!|\!| u^k - u_h^k |\!|\!|^2 \leqslant M_\oplus^k(u_h^k, p_h^k, \tau, \beta) \quad \forall \tau \in H(\mathrm{Div}, \Omega), \ \beta \in \mathbb{R}_+$$

where

$$M_\oplus^k(u_h^k, p_h^k, \tau, \beta) := \int_\Omega H_1(\nu, \mu, \beta) r^2(u_h^k, \tau)\mathrm{d}x + H_2(\beta)\|\sqrt{\nu}^{-1} d(u_h^k, p_h^k, \tau)\|_\Sigma^2$$

and

$$H_1(\nu,\mu,\beta) := \frac{C_\mathrm{F}^2(1+\beta)}{\underline{\nu} + C_\mathrm{F}^2(1+\beta)\mu} \tag{3.1}$$

$$H_2(\beta) := 1 + \beta^{-1} \tag{3.2}$$

$$r(u_h^k, \tau) := f - \mu u_h^k + \mathrm{Div}\,\tau \tag{3.3}$$

$$d(u_h^k, p_h^k, \tau) := \tau - \nu\nabla u_h^k + \mathbb{I} p_h^k. \tag{3.4}$$

Here $\mathbb{I}$ *denotes the unit tensor.*

Proof. By equation (1.5) we have

$$\int_\Omega \Big(\nu \nabla u^k : \nabla w + \mu u^k \cdot w\Big)\,\mathrm{d}x \;=\; \int_\Omega \Big(f \cdot w + p_h^k \operatorname{div} w\Big)\,\mathrm{d}x.$$

We subtract the integral $\int_\Omega \big(\nu \nabla u_h^k : \nabla w + \mu u_h^k \cdot w\big)\,\mathrm{d}x$ from both sides of the above equation, and obtain

$$\begin{aligned}\int_\Omega \nu \nabla (u^k - u_h^k) : \nabla w + \mu (u^k - u_h^k) \cdot w \,\mathrm{d}x& \\ = \int_\Omega \Big((f - \mu u_h^k) \cdot w - \nu \nabla u_h^k : \nabla w + p_h^k \operatorname{div} w\Big)\,\mathrm{d}x.& \qquad (3.5)\end{aligned}$$

By adding (2.10) to the right-hand side of (3.5) we have

$$\begin{aligned}&\int_\Omega \Big(\nu \nabla (u^k - u_h^k) : \nabla w + \mu (u^k - u_h^k) \cdot w\Big)\,\mathrm{d}x \\ &\quad= \int_\Omega \Big((f - \mu u_h^k + \operatorname{Div} \tau) \cdot w + (\tau - \nu \nabla u_h^k + \mathbb{I} p_h^k) : \nabla w\Big)\,\mathrm{d}x \\ &\quad= \int_\Omega \Big(r(u_h^k, \tau) \cdot w + d(u_h^k, p_h^k, \tau) : \nabla w\Big)\,\mathrm{d}x \qquad (3.6)\end{aligned}$$

where we have used the notation (3.3) and (3.4). Note that

$$\begin{aligned}\int_\Omega r \cdot w \,\mathrm{d}x &= \int_\Omega \Big(\sqrt{\mu}^{-1} \alpha r \cdot \sqrt{\mu} w + (1-\alpha) r \cdot w\Big)\,\mathrm{d}x \\ &\leqslant \|\sqrt{\mu}^{-1} \alpha r\| \, \|\sqrt{\mu} w\| + \|(1-\alpha) r\| \, \|w\| \\ &\leqslant \|\sqrt{\mu}^{-1} \alpha r\| \, \|\sqrt{\mu} w\| + C_{\mathrm{F}} \sqrt{\underline{\nu}}^{-1} \|(1-\alpha) r\| \, \|\sqrt{\nu} \nabla w\|_\Sigma \qquad (3.7)\end{aligned}$$

where $0 \leqslant \alpha(x) \leqslant 1$. Also, we have

$$\int_\Omega d : \nabla w \,\mathrm{d}x \;\leqslant\; \|\sqrt{\nu}^{-1} d\|_\Sigma \|\sqrt{\nu} \nabla w\|_\Sigma. \qquad (3.8)$$

By (3.7) and (3.8) the right-hand side of (3.6) becomes

$$\begin{aligned}&\Big(C_{\mathrm{F}} \sqrt{\underline{\nu}}^{-1} \|(1-\alpha) r\| + \|\sqrt{\nu}^{-1} d\|_\Sigma\Big) \|\sqrt{\nu} \nabla w\|_\Sigma + \|\sqrt{\mu}^{-1} \alpha r\| \, \|\sqrt{\mu} w\| \\ &\quad\leqslant \sqrt{\Big(C_{\mathrm{F}} \sqrt{\underline{\nu}}^{-1} \|(1-\alpha) r\| + \|\sqrt{\nu}^{-1} d\|_\Sigma\Big)^2 + \|\sqrt{\mu}^{-1} \alpha r\|^2} \; ||| \, w \, ||| \,. \qquad (3.9)\end{aligned}$$

We set $w = u^k - u_h^k$, use (3.6) and (3.9), and obtain

$$\begin{aligned}||| \, u^k - u_h^k \, |||^2 &\leqslant \Big(C_{\mathrm{F}} \sqrt{\underline{\nu}}^{-1} \|(1-\alpha) r\| + \|\sqrt{\nu}^{-1} d\|_\Sigma\Big)^2 + \|\sqrt{\mu}^{-1} \alpha r\|^2 \\ &\leqslant (1+\beta) C_{\mathrm{F}}^2 \underline{\nu}^{-1} \|(1-\alpha) r\|^2 \\ &\quad + \; (1+\beta^{-1}) \|\sqrt{\nu}^{-1} d\|_\Sigma^2 \; + \; \|\sqrt{\mu}^{-1} \alpha r\|^2. \qquad (3.10)\end{aligned}$$

It is easy to see that the optimal value of α is defined by the relation

$$\alpha = \frac{C_{\mathrm{F}}^2(1+\beta)\mu}{\underline{\nu} + C_{\mathrm{F}}^2(1+\beta)\mu} \tag{3.11}$$

so that (3.10) implies the estimate

$$\begin{aligned} |\!|\!| u^k - u_h^k |\!|\!|^2 &\leqslant \int_\Omega \frac{C_{\mathrm{F}}^2(1+\beta)}{\underline{\nu} + C_{\mathrm{F}}^2(1+\beta)\mu} r^2 \, \mathrm{d}x + (1+\beta^{-1}) \| \sqrt{\nu}^{-1} d \|_\Sigma^2 \\ &= \int_\Omega H_1 r^2 \, \mathrm{d}x + H_2 \| \sqrt{\nu}^{-1} d \|_\Sigma^2 \end{aligned}$$

where we have used the notation (3.1) and (3.2).

Remark 3.1. It is easy to see that the upper bound $M_\oplus^k$ is sharp. Indeed, by setting $\tau = \nu \nabla u^k - \mathbb{I} p_h^k$, and letting β tend to infinity, we get the exact error in the energy norm $|\!|\!| \cdot |\!|\!|$.

A similar estimate can be derived for the problem generated by the augmented Lagrangian L_A.

Theorem 3.2. *Let (u^k, p_h^k) be the exact solution on the iteration k of the Uzawa algorithm. Then, for the solutions calculated by* (1.6)*, and for an approximation $u_h^k \in V_{0h} + u_D$ we have*

$$|\!|\!| u^k - u_h^k |\!|\!|^2 \leqslant]\!]\!] u^k - u_h^k |\!|\!|_\lambda^2 \leqslant M_\oplus^{k,\lambda}(u_h^k, p_h^k, \tau, \beta) \quad \forall \tau \in H(\mathrm{Div}, \Omega), \;\; \beta \in \mathbb{R}_+$$

where

$$M_\oplus^{k,\lambda}(u_h^k, p_h^k, \tau, \beta) := \int_\Omega H_1(\nu, \mu, \beta) r^2(u_h^k, \tau) \mathrm{d}x + H_2(\beta) \| \sqrt{\nu}^{-1} d^\lambda(u_h^k, p_h^k, \tau) \|_\Sigma^2.$$

The quantities H_1, H_2, and r are defined in (3.1)–(3.3)*, and*

$$d^\lambda(u_h^k, p_h^k, \tau) := \tau - \nu \nabla u_h^k + \mathbb{I}(p_h^k - \lambda \operatorname{div} u_h^k). \tag{3.12}$$

Proof. By (1.6), we have

$$\int_\Omega \Big(\nu \nabla u^k : \nabla w + \mu u^k \cdot w + \lambda \operatorname{div} u^k \operatorname{div} w \Big) \mathrm{d}x = \int_\Omega \Big(f \cdot w + p_h^k \operatorname{div} w \Big) \mathrm{d}x.$$

We subtract the integral $\int_\Omega \big(\nu \nabla u_h^k : \nabla w + \mu u_h^k \cdot w + \lambda \operatorname{div} u_h^k \operatorname{div} w \big) \mathrm{d}x$ from both sides of the above equation, and use (2.10), and obtain

$$
\begin{aligned}
&\int_\Omega \Big(\nu\nabla(u^k-u_h^k):\nabla w+\mu(u^k-u_h^k)\cdot w+\lambda\operatorname{div}(u^k-u_h^k)\operatorname{div}w\Big)\,\mathrm{d}x\\
&\quad=\int_\Omega\Big((f-\mu u_h^k)\cdot w-\nu\nabla u_h^k:\nabla w+(p_h^k-\lambda\operatorname{div}u_h^k)\operatorname{div}w\Big)\,\mathrm{d}x\\
&\quad=\int_\Omega\Big((f-\mu u_h^k+\operatorname{Div}\tau)\cdot w+\Big(\tau-\nu\nabla u_h^k+\mathbb{I}(p_h^k-\lambda\operatorname{div}u_h^k)\Big):\nabla w\Big)\,\mathrm{d}x\\
&\quad=\int_\Omega\Big(r(u_h^k,\tau)\cdot w+d^\lambda(u_h^k,p_h^k,\tau):\nabla w\Big)\,\mathrm{d}x \qquad (3.13)
\end{aligned}
$$

where we have used the notation (3.3) and (3.12). By the same arguments as in (3.7) and (3.8), we represent the right-hand side of (3.13) in the form

$$
\begin{aligned}
&\Big(C_{\mathrm{F}}\sqrt{\underline{\nu}}^{-1}\|(1-\alpha)r\|+\|\sqrt{\nu}^{-1}d^\lambda\|_\Sigma\Big)\,\|\sqrt{\nu}\nabla w\|_\Sigma+\|\sqrt{\mu}^{-1}\alpha r\|\;\|\sqrt{\mu}w\|\\
&\quad\leqslant\sqrt{\Big(C_{\mathrm{F}}\sqrt{\underline{\nu}}^{-1}\|(1-\alpha)r\|+\|\sqrt{\nu}^{-1}d^\lambda\|_\Sigma\Big)^2+\|\sqrt{\mu}^{-1}\alpha r\|^2}\;]\!|\!|\, w\,|\!|\![_\lambda \qquad (3.14)
\end{aligned}
$$

since $|\!|\!|\, w\,|\!|\!|\leqslant\,]\!|\!|\, w\,|\!|\![_\lambda$. By choosing $w=u^k-u_h^k$, (3.13) and (3.14) give

$$
\begin{aligned}
]\!|\!|\, u^k-u_h^k\,|\!|\![_\lambda^2 &\leqslant\Big(C_{\mathrm{F}}\sqrt{\underline{\nu}}^{-1}\|(1-\alpha)r\|+\|\sqrt{\nu}^{-1}d^\lambda\|_\Sigma\Big)^2+\|\sqrt{\mu}^{-1}\alpha r\|^2\\
&\leqslant(1+\beta)C_{\mathrm{F}}^2\underline{\nu}^{-1}\|(1-\alpha)r\|^2\\
&\qquad+\ (1+\beta^{-1})\|\sqrt{\nu}^{-1}d^\lambda\|_\Sigma^2\ +\ \|\sqrt{\mu}^{-1}\alpha r\|^2.
\end{aligned}
$$

Again, we see that the optimal value of α is given by the relation (3.11), and obtain

$$
\begin{aligned}
]\!|\!|\, u^k-u_h^k\,|\!|\![_\lambda^2 &\leqslant\int_\Omega\frac{C_{\mathrm{F}}^2(1+\beta)}{\underline{\nu}+C_{\mathrm{F}}^2(1+\beta)\mu}r^2\,\mathrm{d}x+(1+\beta^{-1})\|\sqrt{\nu}^{-1}d^\lambda\|_\Sigma^2\\
&=\int_\Omega H_1r^2\,\mathrm{d}x+H_2\|\sqrt{\nu}^{-1}d^\lambda\|_\Sigma^2
\end{aligned}
$$

where we have used the notation (3.1) and (3.2).

Finally, by using Theorems 2.2, 3.1, and 3.2 we obtain the final result.

Theorem 3.3. *Let u be the exact velocity, (u^k,p_h^k) be the exact solution calculated on the iteration k of the Uzawa algorithm, and $u_h^k\in V_{0h}+u_D$ be an approximation of the velocity calculated on this iteration. For* (1.5) *we have*

$$
|\!|\!|\, u-u_h^k\,|\!|\!|\leqslant\mathbf{M}_\oplus^k(u_h^k,p_h^k,\tau,\beta)\qquad\forall\tau\in H(\operatorname{Div},\Omega),\ \ \beta\in\mathbb{R}_+
$$

and for (1.6) *we have*

$$
|\!|\!|\, u-u_h^k\,|\!|\!|\leqslant\mathbf{M}_\oplus^{k,\lambda}(u_h^k,p_h^k,\tau,\beta)\qquad\forall\tau\in H(\operatorname{Div},\Omega),\ \ \beta\in\mathbb{R}_+
$$

where

$$\mathbf{M}^k_{\oplus}(u^k_h, p^k_h, \tau, \beta) := 2C\|\operatorname{div} u^k_h\| + (2C\sqrt{\underline{\nu}}^{-1} + 1)\sqrt{M^k_{\oplus}(u^k_h, p^k_h, \tau, \beta)}$$
$$\mathbf{M}^{k,\lambda}_{\oplus}(u^k_h, p^k_h, \tau, \beta) := 2C\|\operatorname{div} u^k_h\| + (2C\sqrt{\underline{\nu}}^{-1} + 1)\sqrt{M^{k,\lambda}_{\oplus}(u^k_h, p^k_h, \tau, \beta)}$$

with C defined in (2.6).

Proof. It is clear that

$$|\!|\!| u - u^k_h |\!|\!| \leqslant |\!|\!| u - u^k |\!|\!| + |\!|\!| u^k - u^k_h |\!|\!| .$$

By Theorem 2.2 we have

$$\begin{aligned} |\!|\!| u - u^k_h |\!|\!| &\leqslant 2C\|\operatorname{div} u^k\| + |\!|\!| u^k - u^k_h |\!|\!| \\ &\leqslant 2C\|\operatorname{div} u^k_h\| + 2C\|\operatorname{div}(u^k - u^k_h)\| + |\!|\!| u^k - u^k_h |\!|\!| \\ &\leqslant 2C\|\operatorname{div} u^k_h\| + 2C\sqrt{\underline{\nu}}^{-1}\|\sqrt{\nu}\nabla(u^k - u^k_h)\| + |\!|\!| u^k - u^k_h |\!|\!| \\ &\leqslant 2C\|\operatorname{div} u^k_h\| + (2C\sqrt{\underline{\nu}}^{-1} + 1)\, |\!|\!| u^k - u^k_h |\!|\!| . \end{aligned}$$

Using the upper bounds presented in Theorems 3.1 and 3.2 for the two cases (1.5) and (1.6), respectively, we arrive at the result.

Finally, we note that estimates for the pressure follows from the above derived estimates. The exact pressure in the Uzawa algorithm is calculated by (1.7), i.e.,

$$p^{k+1} = (p^k_h - \rho \operatorname{div} u^k) \in \widetilde{L}_2(\Omega) \tag{3.15}$$

and an approximation of it is calculated within the framework of the selected finite dimensional subspaces, i.e.,

$$p^{k+1}_h = (p^k_h - \rho \operatorname{div} u^k_h) \in Q_h(\Omega). \tag{3.16}$$

Theorem 3.4. *Let* (u^k, p^k_h) *be the exact solution calculated on the iteration k of the Uzawa algorithm, and* $u^k_h \in V_{0h} + u_D$ *be an approximation of the velocity calculated on this iteration. Now, we apply the estimates presented in Theorems* 3.1 *and* 3.2, *and obtain for* (1.5):

$$\|p^{k+1} - p^{k+1}_h\| \leqslant \rho\sqrt{\underline{\nu}}^{-1}\sqrt{M^k_{\oplus}(u^k_h, p^k_h, \tau, \beta)} \quad \forall \tau \in H(\mathrm{Div}, \Omega),\ \beta \in \mathbb{R}_+$$

and for (1.6)

$$\|p^{k+1} - p^{k+1}_h\| \leqslant \rho\sqrt{\underline{\nu}}^{-1}\sqrt{M^{k,\lambda}_{\oplus}(u^k_h, p^k_h, \tau, \beta)} \quad \forall \tau \in H(\mathrm{Div}, \Omega),\ \beta \in \mathbb{R}_+.$$

Proof. Indeed, from (3.15) and (3.16) we find that

$$\begin{aligned}\|p^{k+1} - p_h^{k+1}\| &= \rho\,\|\mathrm{div}(u^k - u_h^k)\| \\ &\leqslant \rho\sqrt{\underline{\nu}}^{-1}\|\sqrt{\nu}\nabla(u^k - u_h^k)\| \\ &\leqslant \rho\sqrt{\underline{\nu}}^{-1}\,|\!|\!|\, u^k - u_h^k \,|\!|\!|\,.\end{aligned}$$

Applying the error bounds presented in Theorems 3.1 and 3.2 completes the proof.

This paper is focused on theoretical analysis of *a posteriori* error bounds for approximations computed by the Uzawa algorithm. However, it is worth adding some comments on the practical applications of the above derived error majorants. The majorants contain the function $\tau \in H(\mathrm{Div}, \Omega)$ and a positive parameter β, which in general can be taken arbitrary. Getting sharp estimates requires a proper selection of them. Finding an optimal β leads to a one-dimensional optimization problem which is easy solvable. The reconstruction of the stress tensor τ based upon computed functions u_h^k and p_h^k provides a reasonable first guess. A better selection can be performed by methods that have been developed and tested for various elliptic problems (see, e.g., [8, 10, 14] and the references cited therein). A systematical study of computational questions in the context of above derived estimates will be exposed in a separate paper, which is now in preparation.

References

1. I. Babuška, The finite element method with Lagrangian multipliers, *Numer. Math.* (1973) **20**, 179–192

2. F. Brezzi, On the existence, uniqueness and approximation of saddle-point problems arising from Lagrangian multipliers, *RAIRO Sér. Rouge Snal. Numér* (1974) **8**(R-2), 129–151.

3. M. Dobrowolski, On the LBB condition in the numerical analysis of the Stokes equations, *Appl. Numer. Math.* (2005) **54**, 314–323.

4. M. Fortin and R. Glowinski, *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems*, North-Holland, New-York, 1983.

5. C. O. Horgan and L. E. Payne, On inequalities of Korn, Friedrichs and Babuška-Aziz. *Arch. Ration. Mech. Anal.* (1982) **82**, No. 2, 165–179.

6. O. A. Ladyzhenskaya, *The Mathematical Theory of Viscous Incompressible Flow*. Gordon and Breach, New York, 1969.

7. O. A. Ladyzhenskaya and V. A. Solonnikov, Some problems of vector analysis and generalized formulations of boundary value problems for the Navier-Stokes equation. *Zap. Nauchn. Semin. LOMI* (1976) **59**, 81–116 (in Russian).

8. P. Neittaanmäki and S. Repin, *Reliable Methods for Computer Simulation. Error Control and A Posteriori Estimates*. Elsevier, Amsterdam, 2004.

9. M. A. Ol'shanskii and E. V. Chizhonkov, On the domain geometry dependence of the LBB condition. *M2AN Math. Model. Numer. Anal.* (2000) **34**, No. 5, 935–951.

10. S. Repin, *A Posteriori Estimates for Partial Differential Equations*. Walter de Gruyter, Berlin, 2008.

11. S. Repin, *A posteriori* estimates for the Stokes problem, *J. Math. Sci.* (2002) **109**, No. 5, 1950–1964.

12. S. Repin and R. Stenberg, *A posteriori* error estimates for the generalized Stokes problem, *J. Math. Sci.* (2007) **142**, No. 1, 1828–1843.

13. R. Temam, *Navier–Stokes Equations. Theory and Numerical Analysis*. North-Holland, New York, 1977.

14. J. Valdman, Minimization of functional majorant in *a posteriori* error analysis based on $H(\mathrm{div})$ multigrid-preconditioned CG method. *Adv. Num. Anal.* 2009, Article ID 164519.

PIII

FUNCTIONAL A POSTERIORI ERROR EQUALITIES FOR CONFORMING MIXED APPROXIMATIONS OF ELLIPTIC PROBLEMS

by

I. Anjam and D. Pauly (2014)

Reports of the Department of Mathematical Information Technology, University of Jyväskylä, Series B. Scientific Computing, No. B. 2/2014, ISBN 978-951-39-5643-1, ISSN 1456-436X

A version available online from arXiv (Cornell University Library)

http://arxiv.org/abs/1403.2560.

Reports of the Department of Mathematical Information Technology
Series B. Scientific Computing
No. B. 2/2014

Functional A Posteriori Error Equalities for Conforming Mixed Approximations of Elliptic Problems

Immanuel Anjam Dirk Pauly

University of Jyväskylä
Department of Mathematical Information Technology
P.O. Box 35 (Agora)
FI–40014 University of Jyväskylä
FINLAND
fax +358 14 260 2771
http://www.mit.jyu.fi/

ISBN 978-951-39-5643-1
ISSN 1456-436X

Functional A Posteriori Error Equalities for Conforming Mixed Approximations of Elliptic Problems

Immanuel Anjam[*] Dirk Pauly[†]

Dedicated to Sergey Igorevich Repin on the occasion of his 60^{th} birthday

Abstract

In this paper we show how to find the *exact error* (not just an estimate of the error) of a conforming mixed approximation by using the functional type a posteriori error estimates in the spirit of Repin [14]. The error is measured in a mixed norm which takes into account both the primal and dual variables. We derive this result for all elliptic partial differential equations of the class

$$\mathrm{A}^* \, \mathrm{A} \; x + x = f,$$

where A is a linear, densely defined and closed (usually a differential) operator and A^* its adjoint. We first derive a special version of our main result by using a simplified reaction-diffusion problem to demonstrate the strong connection to the classical functional a posteriori error estimates of Repin [14]. After this we derive the main result in an abstract setting. Our main result states that in order to obtain the *exact global error* value of a conforming mixed approximation with primal variable x and dual variable y, i.e.,

$$\mathrm{A}^* \, y + x = f, \quad \mathrm{A} \, x = y,$$

one only needs the problem data and the approximation $(\tilde{x}, \tilde{y}) \in D(\mathrm{A}) \times D(\mathrm{A}^*)$ of the exact solution $(x, y) \in D(\mathrm{A}) \times \big(D(\mathrm{A}^*) \cap R(\mathrm{A})\big)$, i.e., the *equality*

$$|x - \tilde{x}|^2 + |\,\mathrm{A}(x - \tilde{x})|^2 + |y - \tilde{y}|^2 + |\,\mathrm{A}^*(y - \tilde{y})|^2 = |f - \tilde{x} - \mathrm{A}^* \, \tilde{y}|^2 + |\tilde{y} - \mathrm{A} \, \tilde{x}|^2$$

holds. There is no need for calculating any auxiliary data. The calculation of the exact error consists of simply calculating two (usually integral) quantities where all the quantities are known after the approximate solution has been obtained by any conforming method guaranteeing $(\tilde{x}, \tilde{y}) \in D(\mathrm{A}) \times D(\mathrm{A}^*)$. We also show some numerical computations to confirm the results.

Key words functional a posteriori error estimate, error equality, elliptic boundary value problem, mixed formulation, combined norm

AMS MSC-Classifications 65N15

[*] Department of Mathematical Information Technology, University of Jyväskylä, PO Box 35 (Agora), FI-40014 University of Jyväskylä, Finland, immanuel.anjam@jyu.fi

[†] Fakultät für Mathematik, Universität Duisburg-Essen, Campus Essen, Thea-Leymann-Str. 9, DE-45141 Essen, Germany, dirk.pauly@uni-due.de

1 Introduction

The results presented in this paper are based on the conception of functional type a posteriori error estimates. These type estimates are valid for any conforming approximation and contain only global constants. We note that estimates for nonconforming approximations are known as well but will not be discussed in this paper. In the case of the class of PDEs studied in this paper, the estimates do not contain even global constants. For a detailed exposition of the theory see the books [14] by Repin and [9] by Repin and Neittaanmäki or for a more computational point of view [8] by Mali, Repin, and Neittaanmäki.

We will measure the error of our approximations in a combined norm, which includes the error of both, the primal and the dual variable. This is especially useful for mixed methods where one calculates an approximation for both the primal and dual variables, see e.g. the book of Brezzi and Fortin [2].

In this paper, we study the linear equation

$$(\mathrm{A}^* \alpha_2 \mathrm{A} + \alpha_1)x = f$$

presented in the mixed form

$$\mathrm{A}^* y + \alpha_1 x = f, \quad \alpha_2 \mathrm{A} = y,$$

where α_1, α_2 are linear and self adjoint topological isomorphisms on two Hilbert spaces H_1 and H_2 and $\mathrm{A} : D(\mathrm{A}) \subset \mathsf{H}_1 \to \mathsf{H}_2$ is a linear, densely defined and closed operator with adjoint operator $\mathrm{A}^* : D(\mathrm{A}^*) \subset \mathsf{H}_2 \to \mathsf{H}_1$. Our main result is Theorem 3.4 and it shortly reads as the *functional a posterior error equality*

$$\begin{aligned} &|x - \tilde{x}|^2_{\mathsf{H}_1,\alpha_1} + |\mathrm{A}(x - \tilde{x})|^2_{\mathsf{H}_2,\alpha_2} + |y - \tilde{y}|^2_{\mathsf{H}_2,\alpha_2^{-1}} + |\mathrm{A}^*(y - \tilde{y})|^2_{\mathsf{H}_1,\alpha_1^{-1}} \\ = &|f - \alpha_1 \tilde{x} - \mathrm{A}^* \tilde{y}|^2_{\mathsf{H}_1,\alpha_1^{-1}} + |\tilde{y} - \alpha_2 \mathrm{A} \tilde{x}|^2_{\mathsf{H}_2,\alpha_2^{-1}} \end{aligned} \tag{1.1}$$

being valid for any conforming mixed approximation $(\tilde{x}, \tilde{y}) \in D(\mathrm{A}) \times D(\mathrm{A}^*)$ of the exact solution $(x, y) \in D(\mathrm{A}) \times D(\mathrm{A}^*)$.

Functional a posteriori error estimates for combined norms were first exposed in the paper [16], where the authors present two-sided estimates bounding the error by the same quantity from below and from above aside from multiplicative constants. Unlike in other estimates, these constants are 1 and $\sqrt{3}$. In [16] the authors studied problems of the type

$$\mathrm{A}^* \alpha \mathrm{A} x = f, \tag{1.2}$$

i.e., the case $\alpha = \alpha_2$, $\alpha_1 = 0$.

The paper is organized as follows. In Section 2 we prove our main results for a simple model problem and show the strong connection to the classical functional a posteriori error estimates. In Section 3 we derive our main results in an abstract Hilbert space setting and in Section 4 we show applications of the general results to several classical problems. Section 5 is devoted to inhomogeneous boundary conditions and finally in Section 6 we present some numerical experiments to confirm our theoretical results.

2 Results for a Model Problem

Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$, be open and without loss of generality connected, so let Ω be a domain with boundary $\Gamma := \partial\Omega$. We emphasize that Ω may be bounded or unbounded, like an exterior domain, or non of both. Moreover, Γ does not need to have any smoothness. We denote by $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathsf{L}^2}$ and $|\,\cdot\,|_{\mathsf{L}^2}$ the inner product and the norm in L^2 for scalar-, vector- and matrix-valued functions. Throughout the paper we will not indicate the dependence on Ω in our notations of the functional spaces. Moreover, we define the usual Sobolev spaces

$$\mathsf{H}^1 := \{\varphi \in \mathsf{L}^2 \mid \nabla\varphi \in \mathsf{L}^2\}, \quad \mathsf{D} := \{\psi \in \mathsf{L}^2 \mid \operatorname{div}\psi \in \mathsf{L}^2\}$$

and as the closure of smooth and compactly supported test functions[1]

$$\mathsf{H}^1_\Gamma := \overline{\mathsf{C}^\infty_\Gamma}^{\mathsf{H}^1}.$$

These are Hilbert spaces equipped with the respective graph norms denoted by $|\,\cdot\,|_{\mathsf{H}^1}$, $|\,\cdot\,|_{\mathsf{D}}$.

Our simple model reaction-diffusion problem reads as follows: Find the potential $u \in \mathsf{H}^1_\Gamma$, i.e., the primal variable, such that

$$-\Delta u + u = -\operatorname{div}\nabla u + u = f, \tag{2.1}$$

where $f \in \mathsf{L}^2$ is the source term. The variational formulation of this problem consists of finding $u \in \mathsf{H}^1_\Gamma$ such that

$$\forall\,\varphi \in \mathsf{H}^1_\Gamma \qquad \langle\nabla u, \nabla\varphi\rangle_{\mathsf{L}^2} + \langle u, \varphi\rangle_{\mathsf{L}^2} = \langle f, \varphi\rangle_{\mathsf{L}^2}. \tag{2.2}$$

The natural energy norm for this problem is $|\,\cdot\,|_{\mathsf{H}^1}$. Of course, by the Lax-Milgram lemma or Riesz' representation theorem (2.2) has a unique solution $u \in \mathsf{H}^1_\Gamma$ satisfying

$$|u|_{\mathsf{H}^1} \leq |f|_{\mathsf{L}^2}.$$

Often, a variable of interest is also the flux, i.e., the dual variable,

$$p := \nabla u \in \mathsf{D},$$

leading to the mixed formulation

$$-\operatorname{div} p + u = f, \quad \nabla u = p.$$

We note that indeed by (2.2) the flux p belongs to D and $\operatorname{div} p = u - f$ holds. Let us further emphasize that even

$$p \in \mathsf{D} \cap \nabla\,\mathsf{H}^1_\Gamma$$

holds, this is, p is also irrotational, has got vanishing tangential trace and is L^2-perpendicular to the so-called Dirichlet fields.

We will understand a pair $(\tilde{u}, \tilde{p}) \in \mathsf{H}^1_\Gamma \times \mathsf{D}$ without further requirements as an approximation of the exact solution pair $(u, p) \in \mathsf{H}^1_\Gamma \times \mathsf{D}$. For the convenience of the reader, we first present the classical functional error upper bounds, frequently called *error majorants*, for the approximations of u and p.

[1] The spaces C^∞_Γ and H^1_Γ are often denoted by $\mathsf{C}^\infty_\circ$ and $\mathsf{H}^1_\circ$.

Theorem 2.1. *For any approximation $\tilde{u} \in \mathsf{H}^1_\Gamma$ of the exact potential u*

$$|u - \tilde{u}|^2_{\mathsf{H}^1} = \min_{\psi \in \mathsf{D}} \mathcal{M}_\nabla(\tilde{u}, \psi) = \mathcal{M}_\nabla(\tilde{u}, p), \tag{2.3}$$

holds, where

$$\mathcal{M}_\nabla(\tilde{u}, \psi) := |f - \tilde{u} + \operatorname{div} \psi|^2_{\mathsf{L}^2} + |\psi - \nabla \tilde{u}|^2_{\mathsf{L}^2}. \tag{2.4}$$

Proof. To derive the upper bound, we subtract $\langle \nabla \tilde{u}, \nabla \varphi \rangle_{\mathsf{L}^2} + \langle \tilde{u}, \varphi \rangle_{\mathsf{L}^2}$ from both sides of the generalized form (2.2), and obtain for all $\varphi \in \mathsf{H}^1_\Gamma$

$$\langle \nabla (u - \tilde{u}), \nabla \varphi \rangle_{\mathsf{L}^2} + \langle u - \tilde{u}, \varphi \rangle_{\mathsf{L}^2} = \langle f - \tilde{u}, \varphi \rangle_{\mathsf{L}^2} - \langle \nabla \tilde{u}, \nabla \varphi \rangle_{\mathsf{L}^2}. \tag{2.5}$$

For an arbitrary function $\psi \in \mathsf{D}$ and any $\varphi \in \mathsf{H}^1_\Gamma$ we have $\langle \operatorname{div} \psi, \varphi \rangle_{\mathsf{L}^2} + \langle \psi, \nabla \varphi \rangle_{\mathsf{L}^2} = 0$. By adding this to the right hand side of (2.5) it becomes

$$\begin{aligned} \langle \nabla (u - \tilde{u}), \nabla \varphi \rangle_{\mathsf{L}^2} + \langle u - \tilde{u}, \varphi \rangle_{\mathsf{L}^2} &= \langle f - \tilde{u} + \operatorname{div} \psi, \varphi \rangle_{\mathsf{L}^2} + \langle \psi - \nabla \tilde{u}, \nabla \varphi \rangle_{\mathsf{L}^2} \\ &\leq |f - \tilde{u} + \operatorname{div} \psi|_{\mathsf{L}^2} |\varphi|_{\mathsf{L}^2} + |\psi - \nabla \tilde{u}|_{\mathsf{L}^2} |\nabla \varphi|_{\mathsf{L}^2} \qquad (2.6) \\ &\leq \mathcal{M}_\nabla(\tilde{u}, \psi)^{1/2} |\varphi|_{\mathsf{H}^1}. \end{aligned}$$

By choosing $\varphi := u - \tilde{u} \in \mathsf{H}^1_\Gamma$ we obtain $|u - \tilde{u}|^2_{\mathsf{H}^1} \leq \mathcal{M}_\nabla(\tilde{u}, \psi)$. Since $p \in \mathsf{D}$, we see that $\mathcal{M}_\nabla(\tilde{u}, p) = |u - \tilde{u}|^2_{\mathsf{H}^1}$. □

As the majorant $\mathcal{M}_\nabla$ is sharp, it immediately provides a technique to obtain approximations for the exact flux p. Minimizing $M_\nabla(\psi) := \mathcal{M}_\nabla(\tilde{u}, \psi)$ with respect to ψ yields by differentiation for all $\psi \in \mathsf{D}$

$$\begin{aligned} 0 \overset{!}{=} M'_\nabla(p)\psi &= 2\langle f - \tilde{u} + \operatorname{div} p, \operatorname{div} \psi \rangle_{\mathsf{L}^2} + 2\langle p - \nabla \tilde{u}, \psi \rangle_{\mathsf{L}^2} \\ &= 2\langle f + \operatorname{div} p, \operatorname{div} \psi \rangle_{\mathsf{L}^2} + 2\langle p, \psi \rangle_{\mathsf{L}^2} \end{aligned}$$

since $\langle \tilde{u}, \operatorname{div} \psi \rangle_{\mathsf{L}^2} = -\langle \nabla \tilde{u}, \psi \rangle_{\mathsf{L}^2}$ because $\tilde{u} \in \mathsf{H}^1_\Gamma$. Hence the following problem occurs: Find $p \in \mathsf{D}$ such that

$$\forall \psi \in \mathsf{D} \qquad \langle \operatorname{div} p, \operatorname{div} \psi \rangle_{\mathsf{L}^2} + \langle p, \psi \rangle_{\mathsf{L}^2} = -\langle f, \operatorname{div} \psi \rangle_{\mathsf{L}^2}. \tag{2.7}$$

Note that $\tilde{u}$ is not present here and the natural energy norm for this problem is $| \cdot |_{\mathsf{D}}$. Once again, by the Lax-Milgram lemma (2.7) has a unique solution $p \in \mathsf{D}$ satisfying

$$|p|_{\mathsf{D}} \leq |f|_{\mathsf{L}^2}.$$

Since $\nabla u \in \mathsf{D}$ solves (2.7), i.e., with (2.1)

$$\langle \operatorname{div} \nabla u, \operatorname{div} \psi \rangle_{\mathsf{L}^2} = \langle u, \operatorname{div} \psi \rangle_{\mathsf{L}^2} - \langle f, \operatorname{div} \psi \rangle_{\mathsf{L}^2} = -\langle \nabla u, \psi \rangle_{\mathsf{L}^2} - \langle f, \operatorname{div} \psi \rangle_{\mathsf{L}^2},$$

we get indeed $p = \nabla u$.

Remark 2.2.

(i) The variational formulation (2.7) for p can also be achieved by testing (2.1) with $\operatorname{div}\psi$ for all $\psi \in \mathsf{D}$ since

$$\begin{aligned}-\langle f, \operatorname{div}\psi\rangle_{\mathsf{L}^2} &= \langle \operatorname{div}\nabla u, \operatorname{div}\psi\rangle_{\mathsf{L}^2} - \langle u, \operatorname{div}\psi\rangle_{\mathsf{L}^2}\\ &= \langle \operatorname{div}\nabla u, \operatorname{div}\psi\rangle_{\mathsf{L}^2} + \langle \nabla u, \psi\rangle_{\mathsf{L}^2} = \langle \operatorname{div} p, \operatorname{div}\psi\rangle_{\mathsf{L}^2} + \langle p, \psi\rangle_{\mathsf{L}^2}.\end{aligned}$$

(ii) By (2.7)

$$p \perp \mathsf{D}_0 := \{v \in \mathsf{D} \mid \operatorname{div} v = 0\}$$

holds. Thus, by the Helmholtz decomposition, i.e., $\mathsf{L}^2 = \overline{\nabla \mathsf{H}^1_\Gamma} \oplus \mathsf{D}_0$, we get $p \in \overline{\nabla \mathsf{H}^1_\Gamma}$. Here, $\perp$ and $\oplus$ denote orthogonality and the orthogonal sum in L^2.

(iii) (2.7) is the dual problem to (2.2) and its strong formulation in duality to (2.1) is

$$-\nabla \operatorname{div} p + p = \nabla f \tag{2.8}$$

with mixed formulation

$$\nabla v + p = \nabla f, \quad -\operatorname{div} p = v.$$

We note that in general $\operatorname{div} p$ does not belong to H^1_Γ, not even to H^1. On the other hand, by (2.7) we see $\operatorname{div} p + f \in \mathsf{H}^1_\Gamma$ with $\nabla(\operatorname{div} p + f) = p$ and the natural Neumann boundary condition $\operatorname{div} p + f = 0$ at Γ appears. Hence f belongs to H^1_Γ, if and only if $v := -\operatorname{div} p \in \mathsf{H}^1_\Gamma$, and $f \in \mathsf{H}^1$, if and only if $v \in \mathsf{H}^1$. In both cases (2.8) holds and moreover for all $\varphi \in \mathsf{H}^1_\Gamma$

$$\langle \nabla v, \nabla\varphi\rangle_{\mathsf{L}^2} + \langle v, \varphi\rangle_{\mathsf{L}^2} = -\langle p, \nabla\varphi\rangle_{\mathsf{L}^2} + \langle v, \varphi\rangle_{\mathsf{L}^2} + \langle \nabla f, \nabla\varphi\rangle_{\mathsf{L}^2} = \langle \nabla f, \nabla\varphi\rangle_{\mathsf{L}^2},$$

thus $v \in \mathsf{H}^1$ solves in the strong sense $-\Delta v + v = -\Delta f$ and $v = f$ at Γ if $\Delta f \in \mathsf{L}^2$.

Theorem 2.3. *For any approximation $\tilde{p} \in \mathsf{D}$ of the exact flux p*

$$|p - \tilde{p}|^2_{\mathsf{D}} = \min_{\varphi \in \mathsf{H}^1_\Gamma} \mathcal{M}_{\mathrm{div}}(\tilde{p}, \varphi) = \mathcal{M}_{\mathrm{div}}(\tilde{p}, u), \tag{2.9}$$

holds, where

$$\mathcal{M}_{\mathrm{div}}(\tilde{p}, \varphi) := |f - \varphi + \operatorname{div}\tilde{p}|^2_{\mathsf{L}^2} + |\tilde{p} - \nabla\varphi|^2_{\mathsf{L}^2}. \tag{2.10}$$

Proof. We add $-\langle \operatorname{div}\tilde{p}, \operatorname{div}\psi\rangle_{\mathsf{L}^2} - \langle \tilde{p}, \psi\rangle_{\mathsf{L}^2}$ to the both sides of the variational formulation (2.7) and obtain for all $\psi \in \mathsf{D}$

$$\langle \operatorname{div}(p - \tilde{p}), \operatorname{div}\psi\rangle_{\mathsf{L}^2} + \langle p - \tilde{p}, \psi\rangle_{\mathsf{L}^2} = -\langle f + \operatorname{div}\tilde{p}, \operatorname{div}\psi\rangle_{\mathsf{L}^2} - \langle \tilde{p}, \psi\rangle_{\mathsf{L}^2}. \tag{2.11}$$

For any $\varphi \in \mathsf{H}^1_\Gamma$ we have again $\langle \nabla\varphi, \psi\rangle_{\mathsf{L}^2} + \langle \varphi, \operatorname{div}\psi\rangle_{\mathsf{L}^2} = 0$. By adding this to the right hand side of (2.11) it becomes

$$\begin{aligned}\langle \operatorname{div}(p-\tilde{p}), \operatorname{div}\psi\rangle_{\mathsf{L}^2} + \langle p-\tilde{p}, \psi\rangle_{\mathsf{L}^2} &= -\langle f-\varphi+\operatorname{div}\tilde{p}, \operatorname{div}\psi\rangle_{\mathsf{L}^2} - \langle \tilde{p}-\nabla\varphi, \psi\rangle_{\mathsf{L}^2} \\ &\leq |f-\varphi+\operatorname{div}\tilde{p}|_{\mathsf{L}^2}|\operatorname{div}\psi|_{\mathsf{L}^2} + |\tilde{p}-\nabla\varphi|_{\mathsf{L}^2}|\psi|_{\mathsf{L}^2} \qquad (2.12) \\ &\leq \mathcal{M}_{\mathrm{div}}(\tilde{p},\varphi)^{1/2}|\psi|_{\mathsf{D}}.\end{aligned}$$

Choosing $\psi = p-\tilde{p} \in \mathsf{D}$ yields $|p-\tilde{p}|^2_{\mathsf{D}} \leq \mathcal{M}_{\mathrm{div}}(\tilde{p},\varphi)$. Finally $\mathcal{M}_{\mathrm{div}}(\tilde{p},u) = |p-\tilde{p}|^2_{\mathsf{D}}$ follows by $u \in \mathsf{H}^1_\Gamma$. □

As before, the sharpness of the majorant $\mathcal{M}_{\mathrm{div}}$ gives us a technique to obtain approximations of the potential u. In fact, global minimization of $M_{\mathrm{div}}(\varphi) := \mathcal{M}_{\mathrm{div}}(\tilde{p},\varphi)$ with respect to φ would lead to the variational formulation (2.2) for finding u, since for all $\varphi \in \mathsf{H}^1_\Gamma$

$$\begin{aligned}0 \overset{!}{=} M'_{\mathrm{div}}(u)\varphi &= -2\langle f-u+\operatorname{div}\tilde{p}, \varphi\rangle_{\mathsf{L}^2} - 2\langle \tilde{p}-\nabla u, \nabla\varphi\rangle_{\mathsf{L}^2} \\ &= 2\langle u-f, \varphi\rangle_{\mathsf{L}^2} + 2\langle \nabla u, \nabla\varphi\rangle_{\mathsf{L}^2}\end{aligned}$$

since $\langle \operatorname{div}\tilde{p}, \varphi\rangle_{\mathsf{L}^2} = -\langle \tilde{p}, \nabla\varphi\rangle_{\mathsf{L}^2}$ by $\tilde{p} \in \mathsf{D}$.

Finally, we note that the functional a posteriori error majorants $\mathcal{M}_\nabla$ and $\mathcal{M}_{\mathrm{div}}$ contain only the problem data, conforming numerical approximations and the free functions ψ and φ.

We define the combined norm for the reaction-diffusion problem in a canonical way as the sum of the energy norms for the potential and the flux:

$$\|(\varphi,\psi)\|^2 := |\varphi|^2_{\mathsf{H}^1} + |\psi|^2_{\mathsf{D}} = |\varphi|^2_{\mathsf{L}^2} + |\nabla\varphi|^2_{\mathsf{L}^2} + |\psi|^2_{\mathsf{L}^2} + |\operatorname{div}\psi|^2_{\mathsf{L}^2}$$

Remark 2.4. We know $|u|_{\mathsf{H}^1} \leq |f|_{\mathsf{L}^2}$ and $|p|_{\mathsf{D}} \leq |f|_{\mathsf{L}^2}$. It is indeed notable that

$$\|(u,p)\| = |f|_{\mathsf{L}^2}$$

holds, which follows immediately by $f = -\operatorname{div}p + u$ and $p = \nabla u$ since

$$|f|^2_{\mathsf{L}^2} = |\operatorname{div}p|^2_{\mathsf{L}^2} + |u|^2_{\mathsf{L}^2} - 2\langle \operatorname{div}p, u\rangle_{\mathsf{L}^2} = |\operatorname{div}p|^2_{\mathsf{L}^2} + |u|^2_{\mathsf{L}^2} + 2|p|^2_{\mathsf{L}^2} = \|(u,p)\|^2.$$

Hence the solution operator

$$L : \mathsf{L}^2 \to \mathsf{H}^1_\Gamma \times \mathsf{D}; f \mapsto (u,p)$$

has norm $|L| = 1$, i.e., L is an isometry.

Our main result for this simple reaction-diffusion problem basically combines Theorems 2.1 and 2.3. However, we outline that the resulting right hand side does not contain u or p anymore and is even an equality.

Theorem 2.5. *For any approximation* $(\tilde{u}, \tilde{p}) \in \mathsf{H}^1_\Gamma \times \mathsf{D}$ *of the exact solution* (u, p)

$$\|(u, p) - (\tilde{u}, \tilde{p})\|^2 = \mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p}) \tag{2.13}$$

and the normalized counterpart

$$\frac{\|(u, p) - (\tilde{u}, \tilde{p})\|^2}{\|(u, p)\|^2} = \frac{\mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p})}{|f|^2_{\mathsf{L}^2}} \tag{2.14}$$

hold, where

$$\mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p}) := \mathcal{M}_\nabla(\tilde{u}, \tilde{p}) = \mathcal{M}_{\mathrm{div}}(\tilde{p}, \tilde{u}) = |f - \tilde{u} + \operatorname{div} \tilde{p}|^2_{\mathsf{L}^2} + |\tilde{p} - \nabla \tilde{u}|^2_{\mathsf{L}^2}. \tag{2.15}$$

The error in the combined norm can thus be exactly computed by quantities we already know: the given problem data f and the conforming approximation $(\tilde{u}, \tilde{p})$.

Proof. Set $\psi = \tilde{p}$ in (2.6) and $\varphi = \tilde{u}$ in (2.12). Then, for any $\varphi \in \mathsf{H}^1_\Gamma$ and any $\psi \in \mathsf{D}$ we have

$$\langle \nabla(u - \tilde{u}), \nabla \varphi \rangle_{\mathsf{L}^2} + \langle u - \tilde{u}, \varphi \rangle_{\mathsf{L}^2} = \langle f - \tilde{u} + \operatorname{div} \tilde{p}, \varphi \rangle_{\mathsf{L}^2} + \langle \tilde{p} - \nabla \tilde{u}, \nabla \varphi \rangle_{\mathsf{L}^2}, \tag{2.16}$$

$$\langle \operatorname{div}(p - \tilde{p}), \operatorname{div} \psi \rangle_{\mathsf{L}^2} + \langle p - \tilde{p}, \psi \rangle_{\mathsf{L}^2} = -\langle f - \tilde{u} + \operatorname{div} \tilde{p}, \operatorname{div} \psi \rangle_{\mathsf{L}^2} - \langle \tilde{p} - \nabla \tilde{u}, \psi \rangle_{\mathsf{L}^2}. \tag{2.17}$$

Adding (2.16) and (2.17) we obtain

$$\begin{aligned} &\langle \nabla(u - \tilde{u}), \nabla \varphi \rangle_{\mathsf{L}^2} + \langle u - \tilde{u}, \varphi \rangle_{\mathsf{L}^2} + \langle \operatorname{div}(p - \tilde{p}), \operatorname{div} \psi \rangle_{\mathsf{L}^2} + \langle p - \tilde{p}, \psi \rangle_{\mathsf{L}^2} \\ &= \langle f - \tilde{u} + \operatorname{div} \tilde{p}, \varphi - \operatorname{div} \psi \rangle_{\mathsf{L}^2} + \langle \tilde{p} - \nabla \tilde{u}, \nabla \varphi - \psi \rangle_{\mathsf{L}^2}. \end{aligned} \tag{2.18}$$

By choosing $\varphi := u - \tilde{u} \in \mathsf{H}^1_\Gamma$ and $\psi := p - \tilde{p} \in \mathsf{D}$, the left hand side of (2.18) turns to the combined norm of the error of the approximation. Since we have

$$\begin{aligned} \varphi - \operatorname{div} \psi &= u - \tilde{u} - \operatorname{div} p + \operatorname{div} \tilde{p} = f - \tilde{u} + \operatorname{div} \tilde{p}, \\ \nabla \varphi - \psi &= \nabla u - \nabla \tilde{u} - p + \tilde{p} = \tilde{p} - \nabla \tilde{u}, \end{aligned}$$

(2.18) becomes (2.13). Putting $\tilde{u} = 0$, $\tilde{p} = 0$ in (2.13) shows $\|(u, p)\| = |f|_{\mathsf{L}^2}$ and thus (2.14). □

Remark 2.6.

(i) We note the similarity of the error majorants in Theorems 2.1, 2.3 and 2.5.

(ii) It is clear that Theorem 2.5 generalizes Theorems 2.1 and 2.3 since these two can be recovered from Theorem 2.5. We just estimate

$$\mathcal{M}_\nabla(\tilde{u}, p) = |u - \tilde{u}|^2_{\mathsf{H}^1} \leq \|(u, p) - (\tilde{u}, \tilde{p})\|^2 = \mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p}) = \mathcal{M}_\nabla(\tilde{u}, \tilde{p})$$

and note that the left hand side does not depend on $\psi := \tilde{p} \in \mathsf{D}$. Analogously we estimate

$$\mathcal{M}_{\mathrm{div}}(\tilde{p}, u) = |p - \tilde{p}|^2_{\mathsf{D}} \leq \|(u, p) - (\tilde{u}, \tilde{p})\|^2 = \mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p}) = \mathcal{M}_{\mathrm{div}}(\tilde{p}, \tilde{u})$$

and note that the left hand side does not depend on $\varphi := \tilde{u} \in \mathsf{H}^1_\Gamma$.

Remark 2.7. There is a simple proof of Theorem 2.5 using just (2.1) and $p = \nabla u$:

$$\begin{aligned}\mathcal{M}_{\mathrm{mix}}(\tilde{u},\tilde{p}) &= |f - \tilde{u} + \operatorname{div}\tilde{p}|^2_{\mathsf{L}^2} + |\tilde{p} - \nabla\tilde{u}|^2_{\mathsf{L}^2}\\ &= |u - \tilde{u} + \operatorname{div}\tilde{p} - \operatorname{div}p|^2_{\mathsf{L}^2} + |\tilde{p} - p + \nabla u - \nabla\tilde{u}|^2_{\mathsf{L}^2}\\ &= |u - \tilde{u}|^2_{\mathsf{L}^2} + |\operatorname{div}(\tilde{p} - p)|^2_{\mathsf{L}^2} + 2\langle u - \tilde{u}, \operatorname{div}(\tilde{p} - p)\rangle_{\mathsf{L}^2}\\ &\qquad + |\tilde{p} - p|^2_{\mathsf{L}^2} + |\nabla(u - \tilde{u})|^2_{\mathsf{L}^2} + 2\langle\tilde{p} - p, \nabla(u - \tilde{u})\rangle_{\mathsf{L}^2}\\ &= \|(u,p) - (\tilde{u},\tilde{p})\|^2\end{aligned}$$

In the last line we have used as before $\langle u - \tilde{u}, \operatorname{div}(\tilde{p} - p)\rangle_{\mathsf{L}^2} = -\langle\nabla(u - \tilde{u}), \tilde{p} - p\rangle_{\mathsf{L}^2}$ since $u - \tilde{u} \in \mathsf{H}^1_\Gamma$. This shows immediately, that Theorem 2.5 extends to more general situations as well. E.g. inhomogeneous boundary conditions can be treated since only $u - \tilde{u} \in \mathsf{H}^1_\Gamma$ is needed.

3 Results for the General Case

In this section we derive our main result in an abstract setting which allows for mixed boundary conditions as well as coefficients for the PDEs. We will prove the main result by using the simple approach presented in Remark 2.7.

Let H_1 and H_2 be two Hilbert spaces with inner products $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathsf{H}_1}$ and $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathsf{H}_2}$, respectively. Moreover, let $\mathrm{A} : D(\mathrm{A}) \subset \mathsf{H}_1 \to \mathsf{H}_2$ be a densely defined and closed linear operator and $\mathrm{A}^* : D(\mathrm{A}^*) \subset \mathsf{H}_2 \to \mathsf{H}_1$ its adjoint. We note $\mathrm{A}^{**} = \bar{\mathrm{A}} = \mathrm{A}$ and

$$\forall\,\varphi \in D(\mathrm{A}) \quad \forall\,\psi \in D(\mathrm{A}^*) \qquad \langle\mathrm{A}\,\varphi, \psi\rangle_{\mathsf{H}_2} = \langle\varphi, \mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1}. \tag{3.1}$$

Equipped with the natural graph norms $D(\mathrm{A})$ and $D(\mathrm{A}^*)$ are Hilbert spaces. Furthermore, we introduce two linear, self adjoint and positive topological isomorphisms $\alpha_1 : \mathsf{H}_1 \to \mathsf{H}_1$ and $\alpha_2 : \mathsf{H}_2 \to \mathsf{H}_2$. Especially we have

$$\exists\,c > 0 \quad \forall\,\varphi \in \mathsf{H}_1 \qquad c^{-1}|\varphi|^2_{\mathsf{H}_1} \leq \langle\alpha_1\varphi, \varphi\rangle_{\mathsf{H}_1} \leq c|\varphi|^2_{\mathsf{H}_1}$$

and the corresponding holds for α_2. For any inner product and corresponding norm we introduce weighted counterparts with sub-index notation. For example, for elements from H_1 we define a new inner product $\langle\,\cdot\,,\,\cdot\,\rangle_{\mathsf{H}_1,\alpha_1} := \langle\alpha_1\,\cdot\,,\,\cdot\,\rangle_{\mathsf{H}_1}$ and a new induced norm $|\cdot|_{\mathsf{H}_1,\alpha_1}$. Using this notation we can define for $\varphi \in D(\mathrm{A})$ and $\psi \in D(\mathrm{A}^*)$ new weighted norms on $D(\mathrm{A})$, $D(\mathrm{A}^*)$ as well as on the product space $D(\mathrm{A}) \times D(\mathrm{A}^*)$ by

$$\begin{aligned}|\varphi|^2_{D(\mathrm{A}),\alpha_1,\alpha_2} &:= |\varphi|^2_{\mathsf{H}_1,\alpha_1} + |\,\mathrm{A}\,\varphi|^2_{\mathsf{H}_2,\alpha_2},\\ |\psi|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}} &:= |\psi|^2_{\mathsf{H}_2,\alpha_2^{-1}} + |\,\mathrm{A}^*\,\psi|^2_{\mathsf{H}_1,\alpha_1^{-1}},\\ \|(\varphi,\psi)\|^2 &:= |\varphi|^2_{D(\mathrm{A}),\alpha_1,\alpha_2} + |\psi|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}}.\end{aligned}$$

Let $f \in \mathsf{H}_1$. By the Lax-Milgram lemma (or by Riesz' representation theorem) we get immediately:

Lemma 3.1. *The (primal) variational problem*

$$\forall\varphi\in D(\mathrm{A})\qquad \langle \mathrm{A}\,x,\mathrm{A}\,\varphi\rangle_{\mathsf{H}_2,\alpha_2}+\langle x,\varphi\rangle_{\mathsf{H}_1,\alpha_1}=\langle f,\varphi\rangle_{\mathsf{H}_1} \tag{3.2}$$

admits a unique solution $x\in D(\mathrm{A})$ satisfying $|x|_{D(\mathrm{A}),\alpha_1,\alpha_2}\leq|f|_{\mathsf{H}_1,\alpha_1^{-1}}$. Also, $y_x:=\alpha_2\,\mathrm{A}\,x$ belongs to $D(\mathrm{A}^*)$ and $\mathrm{A}^*\,y_x=f-\alpha_1x$. Hence, the strong and mixed formulations

$$\mathrm{A}^*\,\alpha_2\,\mathrm{A}\,x+\alpha_1x=f, \tag{3.3}$$

$$\mathrm{A}^*\,y_x+\alpha_1x=f,\quad \alpha_2\,\mathrm{A}\,x=y_x \tag{3.4}$$

hold with $(x,y_x)\in D(\mathrm{A})\times\big(D(\mathrm{A}^*)\times\alpha_2R(\mathrm{A})\big)$.

To get the dual problem, we multiply the first equation of (3.4) by $\mathrm{A}^*\,\psi$ with $\psi\in D(\mathrm{A}^*)$ taking the right weighted scalar product and use $y_x=\alpha_2\,\mathrm{A}\,x\in D(\mathrm{A}^*)$. We obtain

$$\langle\mathrm{A}^*\,y_x,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}+\langle\alpha_1x,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}=\langle f,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}.$$

Since $x\in D(\mathrm{A})$

$$\langle\alpha_1x,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}=\langle x,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1}=\langle\mathrm{A}\,x,\psi\rangle_{\mathsf{H}_2}=\langle y_x,\psi\rangle_{\mathsf{H}_2,\alpha_2^{-1}}$$

holds, we get again by the Lax-Milgram's lemma

Lemma 3.2. *The (dual) variational problem*

$$\forall\psi\in D(\mathrm{A}^*)\qquad\langle\mathrm{A}^*\,y,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}+\langle y,\psi\rangle_{\mathsf{H}_2,\alpha_2^{-1}}=\langle f,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}} \tag{3.5}$$

admits a unique solution $y\in D(\mathrm{A}^*)$ satisfying $|y|_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}}\leq|f|_{\mathsf{H}_1,\alpha_1^{-1}}$. Moreover, $y=y_x$ holds and thus y even belongs to $D(\mathrm{A}^*)\cap\alpha_2R(\mathrm{A})$ with x and y_x from Lemma 3.1. Furthermore, $\alpha_1^{-1}(\mathrm{A}^*\,y-f)\in D(\mathrm{A})$ with $A\alpha_1^{-1}(\mathrm{A}^*\,y-f)=-\alpha_2^{-1}y$.

Proof. We just have to show that $y_x\in D(\mathrm{A}^*)$ solves (3.5). But this follows directly since for all $\psi\in D(\mathrm{A}^*)$

$$\begin{aligned}\langle\mathrm{A}^*\,y_x,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}&=-\langle x,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1}+\langle f,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}\\&=-\langle\mathrm{A}\,x,\psi\rangle_{\mathsf{H}_2}+\langle f,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}=-\langle y_x,\psi\rangle_{\mathsf{H}_2,\alpha_2^{-1}}+\langle f,\mathrm{A}^*\,\psi\rangle_{\mathsf{H}_1,\alpha_1^{-1}}.\end{aligned}$$

Hence $y_x=y$ and $\mathrm{A}^{**}=\mathrm{A}$ completes the proof. □

Remark 3.3. We know $|x|_{D(\mathrm{A}),\alpha_1,\alpha_2}\leq|f|_{\mathsf{H}_1,\alpha_1^{-1}}$ and $|y|_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}}\leq|f|_{\mathsf{H}_1,\alpha_1^{-1}}$. It is indeed notable that

$$\|(x,y)\|=|f|_{\mathsf{H}_1,\alpha_1^{-1}}$$

holds, which follows immediately by $y=\alpha_2\,\mathrm{A}\,x$ and

$$\begin{aligned}|f|^2_{\mathsf{H}_1,\alpha_1^{-1}}&=|\,\mathrm{A}^*\,\alpha_2\,\mathrm{A}\,x+\alpha_1x|^2_{\mathsf{H}_1,\alpha_1^{-1}}\\&=|\,\mathrm{A}^*\,y|^2_{\mathsf{H}_1,\alpha_1^{-1}}+|\alpha_1x|^2_{\mathsf{H}_1,\alpha_1^{-1}}+2\,\underbrace{\langle\mathrm{A}^*\,\alpha_2\,\mathrm{A}\,x,\alpha_1x\rangle_{\mathsf{H}_1,\alpha_1^{-1}}}_{=\langle\mathrm{A}^*\,\alpha_2\,\mathrm{A}\,x,x\rangle_{\mathsf{H}_1}}\\&=|\,\mathrm{A}^*\,y|^2_{\mathsf{H}_1,\alpha_1^{-1}}+|x|^2_{\mathsf{H}_1,\alpha_1}+2\,\underbrace{\langle\alpha_2\,\mathrm{A}\,x,\mathrm{A}\,x\rangle_{\mathsf{H}_2}}_{=|\,\mathrm{A}\,x|^2_{\mathsf{H}_2,\alpha_2}}=\|(x,y)\|^2.\end{aligned}$$

Thus the solution operator

$$L : \mathsf{H}_1 \to D(\mathrm{A}) \times D(\mathrm{A}^*); f \mapsto (x, y)$$

(equipped with the proper weighted norms) has norm $|L| = 1$, i.e., L is an isometry.

By the latter remark the mixed norm on $D(\mathrm{A}) \times D(\mathrm{A}^*)$ yields an isomtery. This motivates to use the mixed norm also for error estimates. As it turns out, we even obtain an error equality. We present our main result of the paper.

Theorem 3.4. *Let $(x, y), (\tilde{x}, \tilde{y}) \in D(\mathrm{A}) \times D(\mathrm{A}^*)$ be the exact solution of* (3.4) *and any conforming approximation, respectively. Then*

$$\|(x, y) - (\tilde{x}, \tilde{y})\|^2 = \mathcal{M}(\tilde{x}, \tilde{y}) \tag{3.6}$$

and the normalized counterpart

$$\frac{\|(x, y) - (\tilde{x}, \tilde{y})\|^2}{\|(x, y)\|^2} = \frac{\mathcal{M}(\tilde{x}, \tilde{y})}{|f|^2_{\mathsf{H}_1,\alpha_1^{-1}}} \tag{3.7}$$

hold, where

$$\mathcal{M}(\tilde{x}, \tilde{y}) := |f - \alpha_1 \tilde{x} - \mathrm{A}^* \tilde{y}|^2_{\mathsf{H}_1,\alpha_1^{-1}} + |\tilde{y} - \alpha_2 \mathrm{A} \tilde{x}|^2_{\mathsf{H}_2,\alpha_2^{-1}}. \tag{3.8}$$

Proof. Using (3.3) and inserting $0 = \alpha_2 \mathrm{A} x - y$ we get by (3.1)

$$\begin{aligned}
\mathcal{M}(\tilde{x}, \tilde{y}) &= |\alpha_1 x - \alpha_1 \tilde{x} + \mathrm{A}^* y - \mathrm{A}^* \tilde{y}|^2_{\mathsf{H}_1,\alpha_1^{-1}} + |\tilde{y} - y + \alpha_2 \mathrm{A} x - \alpha_2 \mathrm{A} \tilde{x}|^2_{\mathsf{H}_2,\alpha_2^{-1}} \\
&= |x - \tilde{x}|^2_{\mathsf{H}_1,\alpha_1} + |\mathrm{A}^*(y - \tilde{y})|^2_{\mathsf{H}_1,\alpha_1^{-1}} + 2\langle \alpha_1(x - \tilde{x}), \mathrm{A}^*(y - \tilde{y})\rangle_{\mathsf{H}_1,\alpha_1^{-1}} \\
&\qquad + |\tilde{y} - y|^2_{\mathsf{H}_2,\alpha_2^{-1}} + |\mathrm{A}(x - \tilde{x})|^2_{\mathsf{H}_2,\alpha_2} + 2\langle \tilde{y} - y, \alpha_2 \mathrm{A}(x - \tilde{x})\rangle_{\mathsf{H}_2,\alpha_2^{-1}} \\
&= |x - \tilde{x}|^2_{D(\mathrm{A}),\alpha_1,\alpha_2} + |y - \tilde{y}|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}} \\
&\qquad + 2\langle x - \tilde{x}, \mathrm{A}^*(y - \tilde{y})\rangle_{\mathsf{H}_1} - 2\langle \mathrm{A}(x - \tilde{x}), y - \tilde{y}\rangle_{\mathsf{H}_2} \\
&= \|(x, y) - (\tilde{x}, \tilde{y})\|^2.
\end{aligned}$$

(3.7) follows by the isometry property in Remark 3.3, completing the proof. □

We note that the isometry property, i.e., $\|(x, y)\| = |f|_{\mathsf{H}_1,\alpha_1^{-1}}$, can be seen by inserting $(\tilde{x}, \tilde{y}) = (0, 0)$ into (3.6) as well.

Remark 3.5. Theorem 3.4 can also be deduced as a special case of the equation [9, (7.2.14)] in the book of Neittaamäki and Repin.

Remark 3.6. Of course, the majorant $\mathcal{M}$ is continuous. Especially we have

$$\begin{aligned}
\mathcal{M}(\tilde{x}, \tilde{y}) &\xrightarrow{\tilde{x} \to x \text{ in } D(\mathrm{A})} |y - \tilde{y}|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}} = \mathcal{M}(x, \tilde{y}), \\
\mathcal{M}(\tilde{x}, \tilde{y}) &\xrightarrow{\tilde{y} \to y \text{ in } D(\mathrm{A}^*)} |x - \tilde{x}|^2_{D(\mathrm{A}),\alpha_1,\alpha_2} = \mathcal{M}(\tilde{x}, y)
\end{aligned}$$

and $\mathcal{M}(\tilde{x}, \tilde{y}) \to \mathcal{M}(x, y) = 0$ if $(\tilde{x}, \tilde{y}) \to (x, y)$ in $D(\mathrm{A}) \times D(\mathrm{A}^*)$. This suggests that the majorant $\mathcal{M}$ can also be used as an error indicator for adaptive computations, even though the equality (3.6) is global.

Corollary 3.7. *Theorem 3.4 provides the well known a posteriori error estimates for the primal and dual problems.*

- **(i)** *For any $\tilde{x} \in D(\mathrm{A})$ it holds $|x - \tilde{x}|^2_{D(\mathrm{A}),\alpha_1,\alpha_2} = \min_{\psi \in D(\mathrm{A}^*)} \mathcal{M}(\tilde{x}, \psi) = \mathcal{M}(\tilde{x}, y)$.*
- **(ii)** *For any $\tilde{y} \in D(\mathrm{A}^*)$ it holds $|y - \tilde{y}|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}} = \min_{\varphi \in D(\mathrm{A})} \mathcal{M}(\varphi, \tilde{y}) = \mathcal{M}(x, \tilde{y})$.*

Proof. We just have to estimate

$$|x - \tilde{x}|^2_{D(\mathrm{A}),\alpha_1,\alpha_2} \leq \|(x, y) - (\tilde{x}, \tilde{y})\|^2 = \mathcal{M}(\tilde{x}, \tilde{y})$$

and note that the left hand side does not depend on $\tilde{y} \in D(\mathrm{A}^*)$. By setting
$\psi := \tilde{y} \in D(\mathrm{A}^*)$ we get

$$|x - \tilde{x}|^2_{D(\mathrm{A}),\alpha_1,\alpha_2} \leq \inf_{\psi \in D(\mathrm{A}^*)} \mathcal{M}(\tilde{x}, \psi).$$

But for $\psi = y \in D(\mathrm{A}^*)$ we see $\mathcal{M}(\tilde{x}, y) = |x - \tilde{x}|^2_{D(\mathrm{A}),\alpha_1,\alpha_2}$, which proves (i). Analogously, we estimate

$$|y - \tilde{y}|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}} \leq \|(x, y) - (\tilde{x}, \tilde{y})\|^2 = \mathcal{M}(\tilde{x}, \tilde{y})$$

and note that the left hand side does not depend on $\tilde{x} \in D(\mathrm{A})$. Setting $\varphi := \tilde{x} \in D(\mathrm{A})$ we get

$$|y - \tilde{y}|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}} \leq \inf_{\varphi \in D(\mathrm{A})} \mathcal{M}(\varphi, \tilde{y}).$$

But for $\varphi = x \in D(\mathrm{A})$ we see $\mathcal{M}(x, \tilde{y}) = |y - \tilde{y}|^2_{D(\mathrm{A}^*),\alpha_1^{-1},\alpha_2^{-1}}$, which shows (ii). □

Remark 3.8.

- **(i)** Since $y \perp_{\alpha_2^{-1}} N(\mathrm{A}^*)$ by (3.5) we get immediately $y \in \alpha_2 \overline{R(\mathrm{A})}$ by the Helmholtz decomposition $\mathsf{H}_2 = N(\mathrm{A}^*) \oplus_{\alpha_2^{-1}} \alpha_2 \overline{R(\mathrm{A})}$.
- **(ii)** If $\alpha_1^{-1} f \in D(\mathrm{A})$ we have $z := \alpha_1^{-1} \mathrm{A}^* y \in D(\mathrm{A})$ and the strong and mixed formulations of (3.5) read

$$\begin{aligned} \mathrm{A}\,\alpha_1^{-1}\,\mathrm{A}^*\,y + \alpha_2^{-1} y &= \mathrm{A}\,\alpha_1^{-1} f, \\ \mathrm{A}\,z + \alpha_2^{-1} y &= \mathrm{A}\,\alpha_1^{-1} f, \quad \alpha_1^{-1}\,\mathrm{A}^*\,y = z. \end{aligned}$$

Then for all $\varphi \in D(\mathrm{A})$ we have

$$\begin{aligned} \langle \mathrm{A}\,z, \mathrm{A}\,\varphi \rangle_{\mathsf{H}_2,\alpha_2} + \langle z, \varphi \rangle_{\mathsf{H}_1,\alpha_1} &= -\langle y, \mathrm{A}\,\varphi \rangle_{\mathsf{H}_2} + \langle z, \varphi \rangle_{\mathsf{H}_1,\alpha_1} + \langle \mathrm{A}\,\alpha_1^{-1} f, A\varphi \rangle_{\mathsf{H}_2,\alpha_2} \\ &= \langle \mathrm{A}\,\alpha_1^{-1} f, A\varphi \rangle_{\mathsf{H}_2,\alpha_2} \end{aligned}$$

and hence $z \in \big(D(\mathrm{A}) \cap \alpha_1^{-1} R(\mathrm{A}^*)\big) \subset D(\mathrm{A})$ is the unique solution of this variational problem. Moreover, we have $\alpha_2(\mathrm{A}\,z - \mathrm{A}\,\alpha_1^{-1} f) \in D(\mathrm{A}^*)$ and also $\mathrm{A}^*\,\alpha_2(\mathrm{A}\,z - \mathrm{A}\,\alpha_1^{-1} f) = -\alpha_1 z$. If $\alpha_2\,\mathrm{A}\,\alpha_1^{-1} f$ belongs to $D(\mathrm{A}^*)$ then this yields $\alpha_2\,\mathrm{A}\,z \in D(\mathrm{A}^*)$ and the strong equation

$$\mathrm{A}^*\,\alpha_2\,\mathrm{A}\,z + \alpha_1 z = \mathrm{A}^*\,\alpha_2\,\mathrm{A}\,\alpha_1^{-1} f.$$

Our error equalities may also be used to compute the radius of the indeterminacy set of solutions in terms of the radius of the indeterminacy set of right hand sides. Often the right hand f of a problem is not known exactly but known to belong to an indeterminacy ball around some known mean data $\hat{f}$. Let us write $f = \hat{f} + f_{\mathsf{osc}}$. Since the solution operator L from Remark 3.3 is an isometry, we have for the solutions $(x, y) = (\hat{x}, \hat{y}) + (x_{\mathsf{osc}}, y_{\mathsf{osc}})$

$$\|(x_{\mathsf{osc}}, y_{\mathsf{osc}})\| = \|L f_{\mathsf{osc}}\| = |f_{\mathsf{osc}}|_{\mathsf{H}_1,\alpha_1^{-1}}.$$

Hence, the solutions belong to a ball of the same radius as the data. In other words, any modeling error is mapped to an error of same size. If the magnitude of the oscillating part f_{osc} is known, we also know the magnitude of variations of the solution set.

3.1 Application to Time Discretization

One main application of our error equalities might be that equations of the type

$$\mathrm{A}^* \alpha_2 \, \mathrm{A} \, x + \alpha_1 x = f \tag{3.9}$$

naturally occur in many types of time discretizations for plenty of linear wave propagation models. A large class of wave propagation models, like electro-magnetics, acoustics or elasticity, have the structure

$$(\partial_t \Lambda^{-1} + \mathrm{M}) \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} g \\ h \end{bmatrix}, \quad \mathrm{M} = \begin{bmatrix} 0 & -\mathrm{A}^* \\ \mathrm{A} & 0 \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

or

$$\partial_t \lambda_1^{-1} x - \mathrm{A}^* \, y = g, \quad \partial_t \lambda_2^{-1} y + \mathrm{A} \, x = h \tag{3.10}$$

with initial condition $(x, y)(0) = (x_0, y_0)$. Often the material is assumed to be time-independent, i.e., Λ does not depend on time. In this case $i\Lambda \, \mathrm{M}$ is selfadjoint in the proper Hilbert spaces and the solution theory follows immediately by the spectral theorem. We note that formally the second order wave equation

$$\left(\partial_t^2 - (\Lambda \, \mathrm{M})^2\right) \begin{bmatrix} x \\ y \end{bmatrix} = (\partial_t - \Lambda \, \mathrm{M})\Lambda \begin{bmatrix} g \\ h \end{bmatrix}, \quad (\Lambda \, \mathrm{M})^2 = \begin{bmatrix} -\lambda_1 \, \mathrm{A}^* \, \lambda_2 \, \mathrm{A} & 0 \\ 0 & -\lambda_2 \, \mathrm{A} \, \lambda_1 \, \mathrm{A}^* \end{bmatrix}$$

holds. A standard implizit time discretization for (3.10) is e.g. the backward Euler scheme, i.e.,

$$\delta_n^{-1} \lambda_1^{-1} (x_n - x_{n-1}) - \mathrm{A}^* \, y_n = g_n, \quad \delta_n^{-1} (y_n - y_{n-1}) + \lambda_2 \, \mathrm{A} \, x_n = \lambda_2 h_n, \quad \delta_n := t_n - t_{n-1}.$$

Hence, we obtain e.g. for x_n

$$\mathrm{A}^* \, \lambda_2 \, \mathrm{A} \, x_n + \delta_n^{-2} \lambda_1^{-1} x_n = f_n := \mathrm{A}^* (\lambda_2 h_n + \delta_n^{-1} y_{n-1}) + \delta_n^{-2} \lambda_1^{-1} x_{n-1} + \delta_n^{-1} g_n$$

provided that $\lambda_2 h_n \in D(\mathrm{A}^*)$. Therefore (3.9) holds for x_n with e.g. $\alpha_1 = \delta_n^{-2} \lambda_1^{-1}$ and $\alpha_2 = \lambda_2$. Of course, a similar equation holds for y_n as well. We note that our arguments extend to 'all' practically used time discretizations.

Functional a posteriori error estimates for wave equations can be found in [15, 12].

4 Applications

We will discuss some standard applications. Let $\Omega \subset \mathbb{R}^d$, $d \geq 1$. Since we want to handle mixed boundary conditions, let us assume for simplicity, that Ω is a bounded or an exterior domain with (compact) Lipschitz continuous boundary Γ. Moreover, let $\Gamma_{\mathtt{D}}$ be an open subset of Γ and $\Gamma_{\mathtt{N}} := \Gamma \setminus \overline{\Gamma_{\mathtt{D}}}$ its complement. We will denote by n the outward unit normal of the boundary. The results presented in this section are direct consequences of Theorem 3.4 and, of course, Lemmas 3.1, 3.2 and Remarks 3.6, 3.3 as well as Corollary 3.7 hold for all special applications.

4.1 Reaction-Diffusion

Find the scalar potential $u \in \mathsf{H}^1$, such that

$$\begin{aligned} -\operatorname{div} \alpha \nabla u + \rho\, u &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\ n \cdot \alpha \nabla u &= 0 && \text{on } \Gamma_{\mathtt{N}}. \end{aligned} \tag{4.1}$$

The quadratic diffusion matrix $\alpha \in \mathsf{L}^\infty$ is symmetric, real valued and uniformly positive definite. The reaction coefficient $\rho \geq \rho_0 > 0$ belongs to L^∞ and the source f to L^2. The dual variable for this problem is the flux $p = \alpha \nabla u \in \mathsf{D}$. We need more Sobolev spaces

$$\mathsf{H}^1_{\Gamma_{\mathtt{D}}} := \overline{\mathsf{C}^\infty_{\Gamma_{\mathtt{D}}}}^{\mathsf{H}^1}, \quad \mathsf{D}_{\Gamma_{\mathtt{N}}} := \overline{\mathsf{C}^\infty_{\Gamma_{\mathtt{N}}}}^{\mathsf{D}}, \quad \mathsf{D}_{\Gamma_{\mathtt{N}},0} := \{\psi \in \mathsf{D}_{\Gamma_{\mathtt{N}}} \mid \operatorname{div} \psi = 0\},$$

where $\mathsf{C}^\infty_{\Gamma_{\mathtt{D}}}$ resp. $\mathsf{C}^\infty_{\Gamma_{\mathtt{N}}}$ are smooth test functions resp. vector fields having supports bounded away from $\Gamma_{\mathtt{D}}$ resp. $\Gamma_{\mathtt{N}}$. In the following we show the relation to the notation of Section 3:

α_1	α_2	A	A^*	H_1	H_2	$D(\mathrm{A})$	$D(\mathrm{A}^*)$
ρ	α	∇	$-\operatorname{div}$	L^2	L^2	$\mathsf{H}^1_{\Gamma_{\mathtt{D}}}$	$\mathsf{D}_{\Gamma_{\mathtt{N}}}$

We note that indeed $D(\mathrm{A}^*) = \mathsf{D}_{\Gamma_{\mathtt{N}}}$ holds for Lipschitz domains, see e.g. [5], which is not trivial at all. The relation (3.1) reads now

$$\forall\, \varphi \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}} \quad \forall\, \psi \in \mathsf{D}_{\Gamma_{\mathtt{N}}} \qquad \langle \nabla \varphi, \psi \rangle_{\mathsf{L}^2} = -\langle \varphi, \operatorname{div} \psi \rangle_{\mathsf{L}^2}.$$

Considering the norms we have

$$\begin{aligned} |u|^2_{\mathsf{H}^1,\rho,\alpha} &= |u|^2_{\mathsf{L}^2,\rho} + |\nabla u|^2_{\mathsf{L}^2,\alpha}, \\ |p|^2_{\mathsf{D},\rho^{-1},\alpha^{-1}} &= |p|^2_{\mathsf{L}^2,\alpha^{-1}} + |\operatorname{div} p|^2_{\mathsf{L}^2,\rho^{-1}}, \\ \|(u,p)\|^2 &= |u|^2_{\mathsf{H}^1,\rho,\alpha} + |p|^2_{\mathsf{D},\rho^{-1},\alpha^{-1}}. \end{aligned}$$

Now (4.1) reads: Find $u \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}}$ with $\alpha \nabla u \in \mathsf{D}_{\Gamma_{\mathtt{N}}}$ such that

$$-\operatorname{div} \alpha \nabla u + \rho\, u = f. \tag{4.2}$$

Equivalently, in mixed formulation we have: Find $(u,p) \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}} \times \mathsf{D}_{\Gamma_{\mathtt{N}}}$ such that

$$-\operatorname{div} p + \rho\, u = f, \quad \alpha \nabla u = p. \tag{4.3}$$

The primal and dual variational problems are: Find $(u,p) \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}} \times \mathsf{D}_{\Gamma_{\mathtt{N}}}$ such that

$$\begin{aligned} &\forall\, \varphi \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}} && \langle \nabla u, \nabla \varphi \rangle_{\mathsf{L}^2,\alpha} + \langle u, \varphi \rangle_{\mathsf{L}^2,\rho} = \langle f, \varphi \rangle_{\mathsf{L}^2}, \\ &\forall\, \psi \in \mathsf{D}_{\Gamma_{\mathtt{N}}} && \langle \operatorname{div} p, \operatorname{div} \psi \rangle_{\mathsf{L}^2,\rho^{-1}} + \langle p, \psi \rangle_{\mathsf{L}^2,\alpha^{-1}} = -\langle f, \operatorname{div} \psi \rangle_{\mathsf{L}^2,\rho^{-1}}. \end{aligned}$$

Theorem 4.1. *Let $(u,p), (\tilde{u},\tilde{p}) \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}} \times \mathsf{D}_{\Gamma_{\mathtt{N}}}$ be the exact solution of* (4.3) *and any approximation, respectively. Then*

$$|\!|(u,p) - (\tilde{u},\tilde{p})|\!|^2 = \mathcal{M}_{\mathrm{rd}}(\tilde{u},\tilde{p}), \qquad \frac{|\!|(u,p) - (\tilde{u},\tilde{p})|\!|^2}{|\!|(u,p)|\!|^2} = \frac{\mathcal{M}_{\mathrm{rd}}(\tilde{u},\tilde{p})}{|f|^2_{\mathsf{L}^2,\rho^{-1}}}$$

hold, where $\mathcal{M}_{\mathrm{rd}}(\tilde{u},\tilde{p}) = |f - \rho\tilde{u} + \operatorname{div}\tilde{p}|^2_{\mathsf{L}^2,\rho^{-1}} + |\tilde{p} - \alpha\nabla\tilde{u}|^2_{\mathsf{L}^2,\alpha^{-1}}$.

Remark 4.2. We note $|u|_{\mathsf{H}^1,\rho,\alpha} \leq |f|_{\mathsf{L}^2,\rho^{-1}}$ and $|p|_{\mathsf{D},\rho^{-1},\alpha^{-1}} \leq |f|_{\mathsf{L}^2,\rho^{-1}}$ and indeed

$$|\!|(u,p)|\!| = |f|_{\mathsf{L}^2,\rho^{-1}}.$$

The solution operator $L : \mathsf{L}^2 \to \mathsf{H}^1_{\Gamma_{\mathtt{D}}} \times \mathsf{D}_{\Gamma_{\mathtt{N}}}; f \mapsto (u,p)$ is an isometry, i.e. $|L| = 1$.

Corollary 4.3. *Theorem 4.1 provides the well known a posteriori error estimates for the primal and dual problems.*

- **(i)** *For any $\tilde{u} \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}}$ it holds $|u - \tilde{u}|^2_{\mathsf{H}^1,\rho,\alpha} = \min\limits_{\psi \in \mathsf{D}_{\Gamma_{\mathtt{N}}}} \mathcal{M}_{\mathrm{rd}}(\tilde{u},\psi) = \mathcal{M}_{\mathrm{rd}}(\tilde{u},p)$.*
- **(ii)** *For any $\tilde{p} \in \mathsf{D}_{\Gamma_{\mathtt{N}}}$ it holds $|p - \tilde{p}|^2_{\mathsf{D},\rho^{-1},\alpha^{-1}} = \min\limits_{\varphi \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}}} \mathcal{M}_{\mathrm{rd}}(\varphi,\tilde{p}) = \mathcal{M}_{\mathrm{rd}}(u,\tilde{p})$.*

Remark 4.4. We have $p = \alpha\nabla u \in \mathsf{D}_{\Gamma_{\mathtt{N}}} \cap \alpha\nabla\,\mathsf{H}^1_{\Gamma_{\mathtt{D}}}$ and u and (u,p) solve (4.2) and (4.3), respectively. Moreover, $\operatorname{div} p + f \in \rho\,\mathsf{H}^1_{\Gamma_{\mathtt{D}}}$ with

$$\nabla\rho^{-1}(\operatorname{div} p + f) = \alpha^{-1} p \in \nabla\,\mathsf{H}^1_{\Gamma_{\mathtt{D}}} = \mathsf{R}_{\Gamma_{\mathtt{D}},0} \cap \mathcal{H}^{\perp}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}}.$$

Hence, for $f \in \rho\,\mathsf{H}^1$ we have $\operatorname{div} p \in \rho\,\mathsf{H}^1$ and therefore the strong and mixed formulations of the dual problem

$$\begin{aligned} -\nabla\rho^{-1}\operatorname{div} p + \alpha^{-1} p &= \nabla\rho^{-1} f && && \text{in } \Omega, \\ \nabla v + \alpha^{-1} p &= \nabla\rho^{-1} f, && -\rho^{-1}\operatorname{div} p = v && \text{in } \Omega \end{aligned}$$

hold, which are completed by the equations

$$\begin{aligned} \operatorname{div} p + f &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\ n \cdot p &= 0 && \text{on } \Gamma_{\mathtt{N}}, \\ \operatorname{rot} \alpha^{-1} p &= 0 && \text{in } \Omega, \\ n \times \alpha^{-1} p &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\ \alpha^{-1} p &\perp \mathcal{H}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}}. \end{aligned}$$

Here the Dirichlet-Neumann fields $\mathcal{H}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}}$ and the space $\mathsf{R}_{\Gamma_{\mathtt{D}},0}$ will be defined in Section 4.2. Of course, $\rho v = f$ on $\Gamma_{\mathtt{D}}$ and by $\rho v \in \operatorname{div}\mathsf{D}_{\Gamma_{\mathtt{N}}}$ we also have $\rho v \perp \mathbb{R}$ if $\Gamma = \Gamma_{\mathtt{N}}$.

For related results and numerical tests for exterior domains see e.g. [10, 7].

4.2 Eddy-Current (3D)

Let $d = 3$. The problem reads: Find the electric field $E \in \mathsf{R}$ such that

$$\begin{aligned} \operatorname{rot} \mu^{-1} \operatorname{rot} E + \epsilon E &= J && \text{in } \Omega, \\ n \times E &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\ n \times \mu^{-1} \operatorname{rot} E &= 0 && \text{on } \Gamma_{\mathtt{N}}, \end{aligned} \tag{4.4}$$

where

$$\mathsf{R} := \{\Phi \in \mathsf{L}^2 \mid \operatorname{rot} \Phi \in \mathsf{L}^2\}, \quad \mathsf{R}_0 := \{\Phi \in \mathsf{R} \mid \operatorname{rot} \Phi = 0\}.$$

We assume that the magnetic permeability μ and the electric permittivity ϵ are symmetric, real valued and uniformly positive definite matrices from L^∞. Of course, the extension to complex valued matrices is straight forward. The electric current J belongs to L^2. The dual variable for this problem is the magnetic field $H = \mu^{-1} \operatorname{rot} E$ which belongs to R. We define the Sobolev spaces

$$\mathsf{R}_{\Gamma_{\mathtt{D}}} := \overline{\mathsf{C}^\infty_{\Gamma_{\mathtt{D}}}}^{\mathsf{R}}, \quad \mathsf{R}_{\Gamma_{\mathtt{D}},0} := \{\Phi \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \mid \operatorname{rot} \Phi = 0\}$$

and analogously $\mathsf{R}_{\Gamma_{\mathtt{N}}}$ and $\mathsf{R}_{\Gamma_{\mathtt{N}},0}$. Moreover, we introduce the co-called Dirichlet-Neumann and Neumann-Dirichlet fields by

$$\begin{aligned} \mathcal{H}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}} &:= \mathsf{R}_{\Gamma_{\mathtt{D}},0} \cap \mathsf{D}_{\Gamma_{\mathtt{N}},0} = \{\Psi \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \cap \mathsf{D}_{\Gamma_{\mathtt{N}}} \mid \operatorname{rot} \Psi = 0 \wedge \operatorname{div} \Psi = 0\}, \\ \mathcal{H}_{\Gamma_{\mathtt{N}},\Gamma_{\mathtt{D}}} &:= \mathsf{R}_{\Gamma_{\mathtt{N}},0} \cap \mathsf{D}_{\Gamma_{\mathtt{D}},0} = \{\Psi \in \mathsf{R}_{\Gamma_{\mathtt{N}}} \cap \mathsf{D}_{\Gamma_{\mathtt{D}}} \mid \operatorname{rot} \Psi = 0 \wedge \operatorname{div} \Psi = 0\}, \end{aligned}$$

respectively. In the following we show the relation to the notation of Section 3:

α_1	α_2	A	A^*	H_1	H_2	$D(\mathrm{A})$	$D(\mathrm{A}^*)$
ϵ	μ^{-1}	rot	rot	L^2	L^2	$\mathsf{R}_{\Gamma_{\mathtt{D}}}$	$\mathsf{R}_{\Gamma_{\mathtt{N}}}$

We note that indeed $D(\mathrm{A}^*) = \mathsf{R}_{\Gamma_{\mathtt{N}}}$ holds for Lipschitz domains, see e.g. [5], which is not trivial at all. The relation (3.1) reads now

$$\forall\, \Phi \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \quad \forall\, \Psi \in \mathsf{R}_{\Gamma_{\mathtt{N}}} \qquad \langle \operatorname{rot} \Phi, \Psi \rangle_{\mathsf{L}^2} = \langle \Phi, \operatorname{rot} \Psi \rangle_{\mathsf{L}^2}.$$

Considering the norms we have

$$\begin{aligned} |E|^2_{\mathsf{R},\epsilon,\mu^{-1}} &= |E|^2_{\mathsf{L}^2,\epsilon} + |\operatorname{rot} E|^2_{\mathsf{L}^2,\mu^{-1}}, \\ |H|^2_{\mathsf{R},\epsilon^{-1},\mu} &= |H|^2_{\mathsf{L}^2,\mu} + |\operatorname{rot} H|^2_{\mathsf{L}^2,\epsilon^{-1}}, \\ \|(E,H)\|^2 &= |E|^2_{\mathsf{R},\epsilon,\mu^{-1}} + |H|^2_{\mathsf{R},\epsilon^{-1},\mu}. \end{aligned}$$

Now (4.4) reads: Find $E \in \mathsf{R}_{\Gamma_{\mathtt{D}}}$ with $\mu^{-1} \operatorname{rot} E \in \mathsf{R}_{\Gamma_{\mathtt{N}}}$ such that

$$\operatorname{rot} \mu^{-1} \operatorname{rot} E + \epsilon E = J.$$

In mixed formulation we have: Find $(E, H) \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{R}_{\Gamma_{\mathtt{N}}}$ such that

$$\operatorname{rot} H + \epsilon E = J, \quad \mu^{-1} \operatorname{rot} E = H.$$

The primal and dual variational problems are: Find $(E, H) \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{R}_{\Gamma_{\mathtt{N}}}$ such that

$$\begin{aligned} &\forall\, \Phi \in \mathsf{R}_{\Gamma_{\mathtt{D}}} && \langle \operatorname{rot} E, \operatorname{rot} \Phi \rangle_{\mathsf{L}^2,\mu^{-1}} + \langle E, \Phi \rangle_{\mathsf{L}^2,\epsilon} = \langle J, \Phi \rangle_{\mathsf{L}^2}, \\ &\forall\, \Psi \in \mathsf{R}_{\Gamma_{\mathtt{N}}} && \langle \operatorname{rot} H, \operatorname{rot} \Psi \rangle_{\mathsf{L}^2,\epsilon^{-1}} + \langle H, \Psi \rangle_{\mathsf{L}^2,\mu} = \langle J, \operatorname{rot} \Psi \rangle_{\mathsf{L}^2,\epsilon^{-1}}. \end{aligned}$$

Theorem 4.5. *For any approximation* $(\tilde{E},\tilde{H}) \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{R}_{\Gamma_{\mathtt{N}}}$

$$\|(E,H)-(\tilde{E},\tilde{H})\|^2 = \mathcal{M}_{\mathrm{ec}}(\tilde{E},\tilde{H}), \qquad \frac{\|(E,H)-(\tilde{E},\tilde{H})\|^2}{\|(E,H)\|^2} = \frac{\mathcal{M}_{\mathrm{ec}}(\tilde{E},\tilde{H})}{|J|^2_{\mathsf{L}^2,\epsilon^{-1}}}$$

hold, where $\mathcal{M}_{\mathrm{ec}}(\tilde{E},\tilde{H}) = |J-\epsilon\tilde{E}-\operatorname{rot}\tilde{H}|^2_{\mathsf{L}^2,\epsilon^{-1}} + |\tilde{H}-\mu^{-1}\operatorname{rot}\tilde{E}|^2_{\mathsf{L}^2,\mu}$.

Remark 4.6. We note $|E|_{\mathsf{R},\epsilon,\mu^{-1}} \leq |J|_{\mathsf{L}^2,\epsilon^{-1}}$ and $|H|_{\mathsf{R},\epsilon^{-1},\mu} \leq |J|_{\mathsf{L}^2,\epsilon^{-1}}$ and indeed

$$\|(E,H)\| = |J|_{\mathsf{L}^2,\epsilon^{-1}}.$$

The solution operator $L : \mathsf{L}^2 \to \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{R}_{\Gamma_{\mathtt{N}}}; f \mapsto (E,H)$ is an isometry, i.e. $|L| = 1$.

Corollary 4.7. *Theorem 4.5 provides the well known a posteriori error estimates for the primal and dual problems.*

- **(i)** *For any* $\tilde{E} \in \mathsf{R}_{\Gamma_{\mathtt{D}}}$ *it holds* $|E-\tilde{E}|^2_{\mathsf{R},\epsilon,\mu^{-1}} = \min_{\Psi \in \mathsf{R}_{\Gamma_{\mathtt{N}}}} \mathcal{M}_{\mathrm{ec}}(\tilde{E},\Psi) = \mathcal{M}_{\mathrm{ec}}(\tilde{E},H)$.
- **(ii)** *For any* $\tilde{H} \in \mathsf{R}_{\Gamma_{\mathtt{N}}}$ *it holds* $|H-\tilde{H}|^2_{\mathsf{R},\epsilon^{-1},\mu} = \min_{\Phi \in \mathsf{R}_{\Gamma_{\mathtt{D}}}} \mathcal{M}_{\mathrm{ec}}(\Phi,\tilde{H}) = \mathcal{M}_{\mathrm{ec}}(E,\tilde{H})$.

Remark 4.8. We have $H = \mu^{-1}\operatorname{rot}E \in \mathsf{R}_{\Gamma_{\mathtt{N}}} \cap \mu^{-1}\operatorname{rot}\mathsf{R}_{\Gamma_{\mathtt{D}}}$ and E and (E,H) solve the strong and mixed formulation, respectively. Moreover, we have $\operatorname{rot}H - J \in \epsilon\,\mathsf{R}_{\Gamma_{\mathtt{D}}}$ with $\operatorname{rot}\epsilon^{-1}(\operatorname{rot}H - J) = -\mu H$ belonging to $\operatorname{rot}\mathsf{R}_{\Gamma_{\mathtt{D}}} = \mathsf{D}_{\Gamma_{\mathtt{D}},0} \cap \mathcal{H}^{\perp}_{\Gamma_{\mathtt{N}},\Gamma_{\mathtt{D}}}$. Hence, for $J \in \epsilon\,\mathsf{R}$ we have $\operatorname{rot}H \in \epsilon\,\mathsf{R}$ and therefore the strong and mixed formulations of the dual problem

$$\begin{aligned} \operatorname{rot}\epsilon^{-1}\operatorname{rot}H + \mu H &= \operatorname{rot}\epsilon^{-1}J && && \text{in } \Omega,\\ \operatorname{rot}D + \mu H &= \operatorname{rot}\epsilon^{-1}J, && \epsilon^{-1}\operatorname{rot}H = D && \text{in } \Omega \end{aligned}$$

hold, which are completed by the equations

$$\begin{aligned} n \times \epsilon^{-1}(\operatorname{rot}H - J) &= 0 && \text{on } \Gamma_{\mathtt{D}},\\ n \times H &= 0 && \text{on } \Gamma_{\mathtt{N}},\\ \operatorname{div}\mu H &= 0 && \text{in } \Omega,\\ n \cdot \mu H &= 0 && \text{on } \Gamma_{\mathtt{D}},\\ \mu H &\perp \mathcal{H}_{\Gamma_{\mathtt{N}},\Gamma_{\mathtt{D}}}. \end{aligned}$$

Of course, $n \times D = n \times \epsilon^{-1}J$ on $\Gamma_{\mathtt{D}}$ and by $\epsilon D \in \operatorname{rot}\mathsf{R}_{\Gamma_{\mathtt{N}}}$ we also have $\operatorname{div}\epsilon D = 0$ in Ω and $n \cdot \epsilon D = 0$ on $\Gamma_{\mathtt{N}}$ as well as $\epsilon D \perp \mathcal{H}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}}$.

Earlier results for eddy current and static Maxwell problems can be found in [1, 11].

4.3 Eddy-Current (2D)

Let $d = 2$. We just indicate the changes compared to the latter section. First, we have to understand the double rot as $\nabla^{\perp} \operatorname{rot}$, where

$$\operatorname{rot} E := \operatorname{div} \mathrm{Q}\, E = \partial_1 E_2 - \partial_2 E_1, \quad \nabla^{\perp} H := \mathrm{Q}\, \nabla H = \begin{bmatrix} \partial_2 H \\ -\partial_1 H \end{bmatrix}, \quad \mathrm{Q} := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$$

and $E \in \mathsf{R}$ is a vector field and $H \in \mathsf{H}^1$ a scalar function. In the literature, the operator $\nabla^{\perp}$ is often called co-gradient or vector rotation $\vec{\operatorname{rot}}$ as well. Also μ is scalar. (4.4) reads: Find the electric field $E \in \mathsf{R}$ such that

$$\begin{aligned} \nabla^{\perp} \mu^{-1} \operatorname{rot} E + \epsilon E &= J && \text{in } \Omega, \\ n \times E &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\ \mu^{-1} \operatorname{rot} E &= 0 && \text{on } \Gamma_{\mathtt{N}}. \end{aligned}$$

We have:

α_1	α_2	A	A^*	H_1	H_2	$D(\mathrm{A})$	$D(\mathrm{A}^*)$
ϵ	μ^{-1}	rot	$\nabla^{\perp}$	L^2	L^2	$\mathsf{R}_{\Gamma_{\mathtt{D}}}$	$\mathsf{H}^1_{\Gamma_{\mathtt{N}}}$

and (3.1) turns to

$$\forall\, \Phi \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \quad \forall\, \psi \in \mathsf{H}^1_{\Gamma_{\mathtt{N}}} \qquad \langle \operatorname{rot} \Phi, \psi \rangle_{\mathsf{L}^2} = \langle \Phi, \nabla^{\perp} \psi \rangle_{\mathsf{L}^2}.$$

The norm for H is

$$|H|^2_{\mathsf{H}^1, \epsilon^{-1}, \mu} = |H|^2_{\mathsf{L}^2, \mu} + |\nabla^{\perp} H|^2_{\mathsf{L}^2, \epsilon^{-1}}.$$

The strong formulation of the problem is: Find $E \in \mathsf{R}_{\Gamma_{\mathtt{D}}}$ with $\mu^{-1} \operatorname{rot} E \in \mathsf{H}^1_{\Gamma_{\mathtt{N}}}$ such that

$$\nabla^{\perp} \mu^{-1} \operatorname{rot} E + \epsilon E = J.$$

The mixed formulation is: Find $(E, H) \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{H}^1_{\Gamma_{\mathtt{N}}}$ such that

$$\nabla^{\perp} H + \epsilon E = J, \quad \mu^{-1} \operatorname{rot} E = H.$$

The primal and dual variational problems are: Find $(E, H) \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{H}^1_{\Gamma_{\mathtt{N}}}$ such that

$$\begin{aligned} &\forall\, \Phi \in \mathsf{R}_{\Gamma_{\mathtt{D}}} && \langle \operatorname{rot} E, \operatorname{rot} \Phi \rangle_{\mathsf{L}^2, \mu^{-1}} + \langle E, \Phi \rangle_{\mathsf{L}^2, \epsilon} = \langle J, \Phi \rangle_{\mathsf{L}^2}, \\ &\forall\, \psi \in \mathsf{H}^1_{\Gamma_{\mathtt{N}}} && \langle \nabla^{\perp} H, \nabla^{\perp} \psi \rangle_{\mathsf{L}^2, \epsilon^{-1}} + \langle H, \psi \rangle_{\mathsf{L}^2, \mu} = \langle J, \nabla^{\perp} \psi \rangle_{\mathsf{L}^2, \epsilon^{-1}}. \end{aligned}$$

Theorem 4.5 reads:

Theorem 4.9. *For any approximation* $(\tilde{E}, \tilde{H}) \in \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{H}^1_{\Gamma_{\mathtt{N}}}$

$$\|(E, H) - (\tilde{E}, \tilde{H})\|^2 = \mathcal{M}_{\mathrm{ec}}(\tilde{E}, \tilde{H}), \qquad \frac{\|(E, H) - (\tilde{E}, \tilde{H})\|^2}{\|(E, H)\|^2} = \frac{\mathcal{M}_{\mathrm{ec}}(\tilde{E}, \tilde{H})}{|J|^2_{\mathsf{L}^2, \epsilon^{-1}}}$$

hold, where $\mathcal{M}_{\mathrm{ec}}(\tilde{E}, \tilde{H}) = |J - \epsilon \tilde{E} - \nabla^{\perp} \tilde{H}|^2_{\mathsf{L}^2, \epsilon^{-1}} + |\tilde{H} - \mu^{-1} \operatorname{rot} \tilde{E}|^2_{\mathsf{L}^2, \mu}$.

Remark 4.10. We note $|E|_{\mathsf{R},\epsilon,\mu^{-1}} \leq |J|_{\mathsf{L}^2,\epsilon^{-1}}$ and $|H|_{\mathsf{H}^1,\epsilon^{-1},\mu} \leq |J|_{\mathsf{L}^2,\epsilon^{-1}}$ and indeed

$$\|(E,H)\| = |J|_{\mathsf{L}^2,\epsilon^{-1}}.$$

The solution operator $L : \mathsf{L}^2 \to \mathsf{R}_{\Gamma_{\mathtt{D}}} \times \mathsf{H}^1_{\Gamma_{\mathtt{N}}}; f \mapsto (E,H)$ is an isometry, i.e. $|L| = 1$.

Corollary 4.11. *Theorem 4.5 provides the well known a posteriori error estimates for the primal and dual problems.*

(i) *For any $\tilde{E} \in \mathsf{R}_{\Gamma_{\mathtt{D}}}$ it holds $|E - \tilde{E}|^2_{\mathsf{R},\epsilon,\mu^{-1}} = \min_{\psi \in \mathsf{H}^1_{\Gamma_{\mathtt{N}}}} \mathcal{M}_{\mathrm{ec}}(\tilde{E},\psi) = \mathcal{M}_{\mathrm{ec}}(\tilde{E},H)$.*

(ii) *For any $\tilde{H} \in \mathsf{H}^1_{\Gamma_{\mathtt{N}}}$ it holds $|H - \tilde{H}|^2_{\mathsf{H}^1,\epsilon^{-1},\mu} = \min_{\Phi \in \mathsf{R}_{\Gamma_{\mathtt{D}}}} \mathcal{M}_{\mathrm{ec}}(\Phi,\tilde{H}) = \mathcal{M}_{\mathrm{ec}}(E,\tilde{H})$.*

Remark 4.12. We have again $H = \mu^{-1} \operatorname{rot} E \in \mathsf{H}^1_{\Gamma_{\mathtt{N}}} \cap \mu^{-1} \operatorname{rot} \mathsf{R}_{\Gamma_{\mathtt{D}}}$ and as in the 3D case E and (E,H) solve the strong and mixed formulation, respectively. Moreover, $\nabla^\perp H - J \in \epsilon\,\mathsf{R}_{\Gamma_{\mathtt{D}}}$ with $\operatorname{rot} \epsilon^{-1}(\nabla^\perp H - J) = -\mu H$. Hence, for $J \in \epsilon\,\mathsf{R}$ we have $\nabla^\perp H \in \epsilon\,\mathsf{R}$ and therefore the strong and mixed formulations of the dual problem

$$\begin{aligned}
\operatorname{rot} \epsilon^{-1} \nabla^\perp H + \mu H &= \operatorname{rot} \epsilon^{-1} J && && \text{in } \Omega, \\
\operatorname{rot} D + \mu H &= \operatorname{rot} \epsilon^{-1} J, && \epsilon^{-1} \nabla^\perp H = D && \text{in } \Omega
\end{aligned}$$

hold, which are completed by the equations

$$\begin{aligned}
n \times \epsilon^{-1}(\nabla^\perp H - J) &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\
H &= 0 && \text{on } \Gamma_{\mathtt{N}}, \\
\mu H &\perp \mathbb{R} \quad (\text{if } \Gamma_{\mathtt{D}} = \Gamma).
\end{aligned}$$

Of course, $n \times D = n \times \epsilon^{-1} J$ on $\Gamma_{\mathtt{D}}$ and by $\epsilon D \in \nabla^\perp \mathsf{H}^1_{\Gamma_{\mathtt{N}}}$ we also have $\operatorname{div} \epsilon D = 0$ in Ω and $n \cdot \epsilon D = 0$ on $\Gamma_{\mathtt{N}}$ as well as $\epsilon D \perp \mathcal{H}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}}$.

4.4 Linear Elasticity

Find the displacement vector field $u \in \mathsf{H}^1$ such that

$$\begin{aligned}
-\operatorname{Div} \Lambda \nabla_{\mathrm{s}} u + \rho\, u &= f && \text{in } \Omega, \\
u &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\
n \cdot \Lambda \nabla_{\mathrm{s}} u &= 0 && \text{on } \Gamma_{\mathtt{N}}.
\end{aligned} \tag{4.5}$$

Here ∇_{s} is the symmetric part of the gradient[2]

$$\nabla_{\mathrm{s}} u := \operatorname{sym} \nabla u = \frac{1}{2}\big(\nabla u + (\nabla u)^\top\big),$$

[2] Here, as usual in elasticity the gradient ∇u is to be understood as the Jacobian of the vector field u.

where $^\top$ denotes the transpose. $\nabla_{\mathrm{s}} u$, often denoted by $\epsilon(u)$, is also called the infinitesimal strain tensor. The fourth order stiffness tensor of elastic moduli $\Lambda \in \mathsf{L}^\infty$, mapping symmetric matrices to symmetric matrices point-wise, and the second order tensor (quadratic matrix) of reaction ρ are assumed to be symmetric, real valued and uniformly positive definite. The vector field f (body force) belongs to L^2 and the dual variable for this problem is the Cauchy stress tensor $\sigma = \Lambda \nabla_{\mathrm{s}} u \in \mathsf{D}$, where the application of Div to σ and the notation $\sigma \in \mathsf{D}$ is to be understood row-wise as the usual divergence div. We note that the first equation can also be written as

$$-\operatorname{Div}_{\mathrm{s}} \Lambda \nabla_{\mathrm{s}} u + \rho\, u = f, \quad \operatorname{Div}_{\mathrm{s}} := \operatorname{Div} \operatorname{sym}.$$

We have:

α_1	α_2	A	A^*	H_1	H_2	$D(\mathrm{A})$	$D(\mathrm{A}^*)$
ρ	Λ	∇_{s}	$-\operatorname{Div}_{\mathrm{s}}$	L^2	L^2	$\mathsf{H}^1_{\Gamma_{\mathsf{D}}}$	$\operatorname{sym}^{-1} \mathsf{D}_{\Gamma_{\mathsf{N}}}$

The notation $\sigma \in \operatorname{sym}^{-1} \mathsf{D}_{\Gamma_{\mathsf{N}}}$ means $\operatorname{sym} \sigma \in \mathsf{D}_{\Gamma_{\mathsf{N}}}$. More precisely, $\psi \in D(\mathrm{A}^*)$ if and only if

$$\forall\, \varphi \in D(\mathrm{A}) = \mathsf{H}^1_{\Gamma_{\mathsf{D}}} \qquad \langle \nabla_{\mathrm{s}} \varphi, \psi \rangle_{\mathsf{L}^2} = \langle \varphi, \mathrm{A}^* \psi \rangle_{\mathsf{L}^2}.$$

Since $\langle \nabla_{\mathrm{s}} \varphi, \psi \rangle_{\mathsf{L}^2} = \langle \nabla \varphi, \operatorname{sym} \psi \rangle_{\mathsf{L}^2}$ we see that this holds if and only if $\operatorname{sym} \psi \in \mathsf{D}_{\Gamma_{\mathsf{N}}}$ and $\mathrm{A}^* \psi = -\operatorname{Div} \operatorname{sym} \psi$. Equation (3.1) turns into

$$\forall\, \varphi \in \mathsf{H}^1_{\Gamma_{\mathsf{D}}} \quad \forall\, \psi \in \operatorname{sym}^{-1} \mathsf{D}_{\Gamma_{\mathsf{N}}} \qquad \langle \nabla_{\mathrm{s}} \varphi, \psi \rangle_{\mathsf{L}^2} = -\langle \varphi, \operatorname{Div}_{\mathrm{s}} \psi \rangle_{\mathsf{L}^2}.$$

For the norms we have

$$\begin{aligned} |u|^2_{\mathsf{H}^1,\rho,\Lambda} &= |u|^2_{\mathsf{L}^2,\rho} + |\nabla_{\mathrm{s}} u|^2_{\mathsf{L}^2,\Lambda}, \\ |\sigma|^2_{\operatorname{sym}^{-1}\mathsf{D},\rho^{-1},\Lambda^{-1}} &= |\sigma|^2_{\mathsf{L}^2,\Lambda^{-1}} + |\operatorname{Div}_{\mathrm{s}} \sigma|^2_{\mathsf{L}^2,\rho^{-1}}, \\ \|(u,\sigma)\|^2 &= |u|^2_{\mathsf{H}^1,\rho,\Lambda} + |\sigma|^2_{\operatorname{sym}^{-1}\mathsf{D},\rho^{-1},\Lambda^{-1}}. \end{aligned}$$

Now (4.5) reads: Find $u \in \mathsf{H}^1_{\Gamma_{\mathsf{D}}}$ with $\operatorname{sym} \Lambda \nabla_{\mathrm{s}} u = \Lambda \nabla_{\mathrm{s}} u \in \mathsf{D}_{\Gamma_{\mathsf{N}}}$ such that

$$-\operatorname{Div} \Lambda \nabla_{\mathrm{s}} u + \rho\, u = f.$$

In mixed formulation we have: Find $(u,\sigma) \in \mathsf{H}^1_{\Gamma_{\mathsf{D}}} \times \mathsf{D}_{\Gamma_{\mathsf{N}}}$ such that

$$-\operatorname{Div} \sigma + \rho\, u = f, \quad \Lambda \nabla_{\mathrm{s}} u = \sigma.$$

Note that then σ is automatically symmetric. The primal and dual variational problems are: Find $(u,\sigma) \in \mathsf{H}^1_{\Gamma_{\mathsf{D}}} \times \operatorname{sym}^{-1} \mathsf{D}_{\Gamma_{\mathsf{N}}}$ such that

$$\begin{aligned} &\forall\, \varphi \in \mathsf{H}^1_{\Gamma_{\mathsf{D}}} && \langle \nabla_{\mathrm{s}} u, \nabla_{\mathrm{s}} \varphi \rangle_{\mathsf{L}^2,\Lambda} + \langle u, \varphi \rangle_{\mathsf{L}^2,\rho} = \langle f, \varphi \rangle_{\mathsf{L}^2}, \\ &\forall\, \psi \in \operatorname{sym}^{-1} \mathsf{D}_{\Gamma_{\mathsf{N}}} && \langle \operatorname{Div}_{\mathrm{s}} \sigma, \operatorname{Div}_{\mathrm{s}} \psi \rangle_{\mathsf{L}^2,\rho^{-1}} + \langle \sigma, \psi \rangle_{\mathsf{L}^2,\Lambda^{-1}} = -\langle f, \operatorname{Div}_{\mathrm{s}} \psi \rangle_{\mathsf{L}^2,\rho^{-1}}. \end{aligned}$$

Since $\sigma \in \mathsf{D}_{\Gamma_{\mathsf{N}}}$ must be symmetric, we can formulate the dual problem also as

$$\forall\, \psi \in \mathsf{D}_{\Gamma_{\mathsf{N}}},\ \psi \text{ symmetric} \qquad \langle \operatorname{Div} \sigma, \operatorname{Div} \psi \rangle_{\mathsf{L}^2,\rho^{-1}} + \langle \sigma, \psi \rangle_{\mathsf{L}^2,\Lambda^{-1}} = -\langle f, \operatorname{Div} \psi \rangle_{\mathsf{L}^2,\rho^{-1}}.$$

Then, the norms reduce to

$$\|(u,\sigma)\|^2 = |u|^2_{\mathsf{H}^1,\rho,\Lambda} + |\sigma|^2_{\mathsf{D},\rho^{-1},\Lambda^{-1}}, \quad |\sigma|^2_{\mathsf{D},\rho^{-1},\Lambda^{-1}} = |\sigma|^2_{\mathsf{L}^2,\Lambda^{-1}} + |\operatorname{Div} \sigma|^2_{\mathsf{L}^2,\rho^{-1}}.$$

Theorem 4.13. *For any approximation* $(\tilde{u}, \tilde{\sigma}) \in \mathsf{H}^1_{\Gamma_\mathtt{D}} \times \operatorname{sym}^{-1} \mathsf{D}_{\Gamma_\mathtt{N}}$

$$\|(u, \sigma) - (\tilde{u}, \tilde{\sigma})\|^2 = \mathcal{M}_{\mathrm{le}}(\tilde{u}, \tilde{\sigma}), \qquad \frac{\|(u, \sigma) - (\tilde{u}, \tilde{\sigma})\|^2}{\|(u, \sigma)\|^2} = \frac{\mathcal{M}_{\mathrm{le}}(\tilde{u}, \tilde{\sigma})}{|f|^2_{\mathsf{L}^2, \rho^{-1}}} \tag{4.6}$$

hold, where $\mathcal{M}_{\mathrm{le}}(\tilde{u}, \tilde{\sigma}) = |f - \rho\tilde{u} + \operatorname{Div}_{\mathrm{s}} \tilde{\sigma}|^2_{\mathsf{L}^2, \rho^{-1}} + |\tilde{\sigma} - \Lambda \nabla_{\mathrm{s}} \tilde{u}|^2_{\mathsf{L}^2, \Lambda^{-1}}$. *Moreover, since* σ *is automatically symmetric we have* (4.6) *for all* $(\tilde{u}, \tilde{\sigma}) \in \mathsf{H}^1_{\Gamma_\mathtt{D}} \times \mathsf{D}_{\Gamma_\mathtt{N}}$ *with* $\tilde{\sigma}$ *symmetric and the right hand side simplifies to* $\mathcal{M}_{\mathrm{le}}(\tilde{u}, \tilde{\sigma}) = |f - \rho\tilde{u} + \operatorname{Div} \tilde{\sigma}|^2_{\mathsf{L}^2, \rho^{-1}} + |\tilde{\sigma} - \Lambda \nabla_{\mathrm{s}} \tilde{u}|^2_{\mathsf{L}^2, \Lambda^{-1}}$.

Remark 4.14. We note $|u|_{\mathsf{H}^1, \rho, \Lambda} \leq |f|_{\mathsf{L}^2, \rho^{-1}}$ and $|\sigma|_{\mathsf{D}, \rho^{-1}, \Lambda^{-1}} \leq |f|_{\mathsf{L}^2, \rho^{-1}}$ and indeed

$$\|(u, \sigma)\| = |f|_{\mathsf{L}^2, \rho^{-1}}.$$

The solution operator $L : \mathsf{L}^2 \to \mathsf{H}^1_{\Gamma_\mathtt{D}} \times \mathsf{D}_{\Gamma_\mathtt{N}}; f \mapsto (u, \sigma)$ is an isometry, i.e. $|L| = 1$.

Corollary 4.15. *Theorem 4.13 provides the well known a posteriori error estimates for the primal and dual problems.*

(i) *For any* $\tilde{u} \in \mathsf{H}^1_{\Gamma_\mathtt{D}}$ *it holds* $|u - \tilde{u}|^2_{\mathsf{H}^1, \rho, \alpha} = \min_{\psi \in \operatorname{sym}^{-1} \mathsf{D}_{\Gamma_\mathtt{N}}} \mathcal{M}_{\mathrm{le}}(\tilde{u}, \psi) = \mathcal{M}_{\mathrm{le}}(\tilde{u}, \sigma)$.

(ii) *For any* $\tilde{\sigma} \in \operatorname{sym}^{-1} \mathsf{D}_{\Gamma_\mathtt{N}}$ *it holds* $|\sigma - \tilde{\sigma}|^2_{\operatorname{sym}^{-1} \mathsf{D}, \rho^{-1}, \alpha^{-1}} = \min_{\varphi \in \mathsf{H}^1_{\Gamma_\mathtt{D}}} \mathcal{M}_{\mathrm{le}}(\varphi, \tilde{\sigma}) = \mathcal{M}_{\mathrm{le}}(u, \tilde{\sigma})$.

If $\tilde{\sigma}$ *and* ψ *are already symmetric we can skip the* sym^{-1} *and replace* $\operatorname{Div}_{\mathrm{s}}$ *by* Div.

Remark 4.16. We have $\sigma = \Lambda \nabla_{\mathrm{s}} u \in \mathsf{D}_{\Gamma_\mathtt{N}} \cap \Lambda \nabla_{\mathrm{s}} \mathsf{H}^1_{\Gamma_\mathtt{D}}$ is symmetric with $\operatorname{Div}_{\mathrm{s}} \sigma = \operatorname{Div} \sigma$ and u and (u, σ) solve the strong and mixed formulation, respectively. Moreover, $\operatorname{Div} \sigma + f \in \rho \mathsf{H}^1_{\Gamma_\mathtt{D}}$ with $\nabla_{\mathrm{s}} \rho^{-1} (\operatorname{Div} \sigma + f) = \Lambda^{-1} \sigma \in \nabla_{\mathrm{s}} \mathsf{H}^1_{\Gamma_\mathtt{D}}$. Hence, for $f \in \rho \mathsf{H}^1$ we have $\operatorname{Div} \sigma \in \rho \mathsf{H}^1$ and therefore strong and mixed formulations of the dual problem hold, i.e.,

$$\begin{aligned} -\nabla_{\mathrm{s}} \rho^{-1} \operatorname{Div} \sigma + \Lambda^{-1} \sigma &= \nabla_{\mathrm{s}} \rho^{-1} f && && \text{in } \Omega, \\ \nabla_{\mathrm{s}} v + \Lambda^{-1} \sigma &= \nabla_{\mathrm{s}} \rho^{-1} f, && -\rho^{-1} \operatorname{Div} \sigma = v && \text{in } \Omega. \end{aligned}$$

4.5 Generalized Reaction-Diffusion, Linear Accoustics and Eddy-Current

Let Ω be a d-dimensional smooth Riemannian manifold with compact Lipschitz boundary Γ. If Ω is unbounded, we assume that outside of some compact set, Ω is isomorphic to the exterior unit domain $\{x \in \mathbb{R}^d \mid |x| > 1\}$. Moreover, let $\Gamma_\mathtt{D}$ be an open subset of Γ and $\Gamma_\mathtt{N} := \Gamma \setminus \overline{\Gamma_\mathtt{D}}$ its complement. The problem reads: For $f \in \mathsf{L}^{2,q}$ find the differential form potential (q-form) $u \in \mathsf{D}^q$, such that

$$\begin{aligned} -\delta \, \alpha \, \mathrm{d} \, u + \rho \, u &= f && \text{in } \Omega, \\ \tau_{\Gamma_\mathtt{D}} u &= 0 && \text{on } \Gamma_\mathtt{D}, \\ \nu_{\Gamma_\mathtt{N}} \alpha \, \mathrm{d} \, u &= 0 && \text{on } \Gamma_\mathtt{N}. \end{aligned} \tag{4.7}$$

Here, d denotes exterior derivative, $\delta = \pm * \mathrm{d} *$ the co-derivative and $\tau_{\Gamma_{\mathtt{D}}}$ resp. $\nu_{\Gamma_{\mathtt{N}}}$ the restrictions of the tangential resp. normal traces τ_Γ resp. ν_Γ to the proper subspaces. We also introduce the Sobolev spaces

$$\mathsf{D}^q := \{\varphi \in \mathsf{L}^{2,q} \mid \mathrm{d}\,\varphi \in \mathsf{L}^{2,q+1}\}, \quad \Delta^q := \{\psi \in \mathsf{L}^{2,q} \mid \delta\,\psi \in \mathsf{L}^{2,q-1}\}$$

and $\mathsf{D}^q_{\Gamma_{\mathtt{D}}} := \overline{\mathsf{C}^{\infty,q}_{\Gamma_{\mathtt{D}}}}^{\mathsf{D}^q}$, $\Delta^q_{\Gamma_{\mathtt{N}}} := \overline{\mathsf{C}^{\infty,q}_{\Gamma_{\mathtt{N}}}}^{\Delta^q}$, where $\mathsf{C}^{\infty,q}_{\Gamma_{\mathtt{D}}}$ resp. $\mathsf{C}^{\infty,q}_{\Gamma_{\mathtt{N}}}$ are smooth test q-forms having supports bounded away from $\Gamma_{\mathtt{D}}$ resp. $\Gamma_{\mathtt{N}}$. Moreover, $\mathsf{L}^{2,q}$ denotes the Lebesgue space of all square integrable q-forms on Ω equipped with the inner or scalar product

$$\langle u, \varphi\rangle_{\mathsf{L}^{2,q}} := \int_\Omega u \wedge *\varphi$$

and corresponding norm $|\cdot|_{\mathsf{L}^{2,q}}$. Of course, D^q and Δ^q are equipped with the respective graph norms, making them Hilbert spaces. Finally, ρ and α denote linear, symmetric, real valued, bounded and uniformly positive definite transformations on q- resp. $(q+1)$-forms. It is again straight forward to discuss complex valued transformations. We also need the spaces

$$\mathsf{D}^q_0 := \{\varphi \in \mathsf{D}^q \mid \mathrm{d}\,\varphi = 0\}, \quad \mathsf{D}^q_{\Gamma_{\mathtt{D}},0} := \{\varphi \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}} \mid \mathrm{d}\,\varphi = 0\}$$

and the corresponding spaces for the co-derivative as well as the space of harmonic Dirichlet-Neumann forms

$$\mathcal{H}^q_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}} := \mathsf{D}^q_{\Gamma_{\mathtt{D}},0} \cap \Delta^q_{\Gamma_{\mathtt{N}},0}\,.$$

The dual variable for this problem is the 'flux' $p = \alpha\,\mathrm{d}\,u \in \Delta^{q+1}$. In the following we show the relation to the notations of Section 3:

α_1	α_2	A	A^*	H_1	H_2	$D(\mathrm{A})$	$D(\mathrm{A}^*)$
ρ	α	d	$-\,\delta$	$\mathsf{L}^{2,q}$	$\mathsf{L}^{2,q+1}$	$\mathsf{D}^q_{\Gamma_{\mathtt{D}}}$	$\Delta^{q+1}_{\Gamma_{\mathtt{N}}}$

Also here indeed $D(\mathrm{A}^*) = \Delta^{q+1}_{\Gamma_{\mathtt{N}}}$ holds, see e.g. [3, 4, 6]. The relation (3.1) turns into

$$\forall\,\varphi \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}} \quad \forall\,\psi \in \Delta^{q+1}_{\Gamma_{\mathtt{N}}} \qquad \langle \mathrm{d}\,\varphi, \psi\rangle_{\mathsf{L}^{2,q+1}} = -\langle \varphi, \delta\,\psi\rangle_{\mathsf{L}^{2,q}}.$$

Considering the norms we have

$$\begin{aligned} |u|^2_{\mathsf{D}^q,\rho,\alpha} &= |u|^2_{\mathsf{L}^{2,q},\rho} + |\,\mathrm{d}\,u|^2_{\mathsf{L}^{2,q+1},\alpha}, \\ |p|^2_{\Delta^{q+1},\rho^{-1},\alpha^{-1}} &= |p|^2_{\mathsf{L}^{2,q+1},\alpha^{-1}} + |\,\delta\,p|^2_{\mathsf{L}^{2,q},\rho^{-1}}, \\ \|(u,p)\|^2 &= |u|^2_{\mathsf{D}^q,\rho,\alpha} + |p|^2_{\Delta^{q+1},\rho^{-1},\alpha^{-1}}. \end{aligned}$$

Now (4.7) reads: Find $u \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}}$ with $\alpha\,\mathrm{d}\,u \in \Delta^{q+1}_{\Gamma_{\mathtt{N}}}$ such that

$$-\,\delta\,\alpha\,\mathrm{d}\,u + \rho\,u = f.$$

In mixed formulation we have: Find $(u,p) \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}} \times \Delta^{q+1}_{\Gamma_{\mathtt{N}}}$ such that

$$-\,\delta\,p + \rho\,u = f, \quad \alpha\,\mathrm{d}\,u = p.$$

The primal and dual variational problems are: Find $(u,p) \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}} \times \Delta^{q+1}_{\Gamma_{\mathtt{N}}}$ such that

$$\begin{aligned} &\forall\, \varphi \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}} && \langle \mathrm{d}\, u, \mathrm{d}\, \varphi\rangle_{\mathsf{L}^{2,q+1},\alpha} + \langle u, \varphi\rangle_{\mathsf{L}^{2,q},\rho} = \langle f, \varphi\rangle_{\mathsf{L}^{2,q}}, \\ &\forall\, \psi \in \Delta^{q+1}_{\Gamma_{\mathtt{N}}} && \langle \delta\, p, \delta\, \psi\rangle_{\mathsf{L}^{2,q},\rho^{-1}} + \langle p, \psi\rangle_{\mathsf{L}^{2,q+1},\alpha^{-1}} = -\langle f, \delta\, \psi\rangle_{\mathsf{L}^{2,q},\rho^{-1}}. \end{aligned}$$

Theorem 4.17. *For any approximation* $(\tilde{u},\tilde{p}) \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}} \times \Delta^{q+1}_{\Gamma_{\mathtt{N}}}$

$$\|(u,p)-(\tilde{u},\tilde{p})\|^2 = \mathcal{M}_{\text{diff}}(\tilde{u},\tilde{p}), \qquad \frac{\|(u,p)-(\tilde{u},\tilde{p})\|^2}{\|(u,p)\|^2} = \frac{\mathcal{M}_{\text{diff}}(\tilde{u},\tilde{p})}{|f|^2_{\mathsf{L}^{2,q},\rho^{-1}}}$$

hold, where $\mathcal{M}_{\text{diff}}(\tilde{u},\tilde{p}) = |f - \rho\tilde{u} + \delta\,\tilde{p}|^2_{\mathsf{L}^{2,q},\rho^{-1}} + |\tilde{p} - \alpha\,\mathrm{d}\,\tilde{u}|^2_{\mathsf{L}^{2,q+1},\alpha^{-1}}$.

Remark 4.18. We note $|u|_{\mathsf{D}^q,\rho,\alpha} \leq |f|_{\mathsf{L}^{2,q},\rho^{-1}}$ and $|p|_{\Delta^{q+1},\rho^{-1},\alpha^{-1}} \leq |f|_{\mathsf{L}^{2,q},\rho^{-1}}$ and indeed

$$\|(u,p)\| = |f|_{\mathsf{L}^{2,q},\rho^{-1}}.$$

The solution operator $L : \mathsf{L}^{2,q} \to \mathsf{D}^q_{\Gamma_{\mathtt{D}}} \times \Delta^{q+1}_{\Gamma_{\mathtt{N}}}; f \mapsto (u,p)$ is an isometry, i.e. $|L| = 1$.

Corollary 4.19. *Theorem 4.17 provides the a posteriori error estimates for the primal and dual problems.*

- **(i)** *For any* $\tilde{u} \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}}$ *it holds* $|u-\tilde{u}|^2_{\mathsf{D}^q,\rho,\alpha} = \min\limits_{\psi \in \Delta^{q+1}_{\Gamma_{\mathtt{N}}}} \mathcal{M}_{\text{diff}}(\tilde{u},\psi) = \mathcal{M}_{\text{diff}}(\tilde{u},p)$.
- **(ii)** *For any* $\tilde{p} \in \Delta^{q+1}_{\Gamma_{\mathtt{N}}}$ *it holds* $|p-\tilde{p}|^2_{\Delta^{q+1},\rho^{-1},\alpha^{-1}} = \min\limits_{\varphi \in \mathsf{D}^q_{\Gamma_{\mathtt{D}}}} \mathcal{M}_{\text{diff}}(\varphi,\tilde{p}) = \mathcal{M}_{\text{diff}}(u,\tilde{p})$.

We note that for $q=0$ we get back the reaction-diffusion problem from Section 4.1 and for $d=3$ or $d=2$ and $q=1$ we obtain the eddy-current problems from Sections 4.2 and 4.3, identifying $\Omega \subset \mathbb{R}^d$ with a proper domain and 0-forms with functions and 1- and 2-forms with vector fields by Riesz' representation theorem and Hodge's star operator.

Remark 4.20. It holds $p = \alpha\,\mathrm{d}\,u \in \Delta^{q+1}_{\Gamma_{\mathtt{N}}} \cap \alpha\,\mathrm{d}\,\mathsf{D}^q_{\Gamma_{\mathtt{D}}}$ and u and (u,p) solve the strong and mixed formulations, respectively. Moreover, $\delta\,p + f$ belongs to $\rho\,\mathsf{D}^q_{\Gamma_{\mathtt{D}}}$ and we see immediately $\mathrm{d}\,\rho^{-1}(\delta\,p + f) = \alpha^{-1}p \in \mathrm{d}\,\mathsf{D}^q_{\Gamma_{\mathtt{D}}} = \mathsf{D}^{q+1}_{\Gamma_{\mathtt{D}},0} \cap (\mathcal{H}^{q+1}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}})^{\perp}$. Hence, for $f \in \rho\,\mathsf{D}^q$ we have $\delta\,p \in \rho\,\mathsf{D}^q$ and therefore the strong and mixed formulations of the dual problem

$$\begin{aligned} -\,\mathrm{d}\,\rho^{-1}\,\delta\,p + \alpha^{-1}p &= \mathrm{d}\,\rho^{-1}f && && \text{in } \Omega, \\ \mathrm{d}\,v + \alpha^{-1}p &= \mathrm{d}\,\rho^{-1}f, && -\rho^{-1}\,\delta\,p = v && \text{in } \Omega \end{aligned}$$

hold, which are completed by the equations

$$\begin{aligned} \tau_{\Gamma_{\mathtt{D}}}\rho^{-1}(\delta\,p + f) &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\ \tau_{\Gamma_{\mathtt{N}}}p &= 0 && \text{on } \Gamma_{\mathtt{N}}, \\ \mathrm{d}\,\alpha^{-1}p &= 0 && \text{in } \Omega, \\ \tau_{\Gamma_{\mathtt{D}}}\alpha^{-1}p &= 0 && \text{on } \Gamma_{\mathtt{D}}, \\ \alpha^{-1}p &\perp \mathcal{H}^{q+1}_{\Gamma_{\mathtt{D}},\Gamma_{\mathtt{N}}}. \end{aligned}$$

There are also more equations for v following from $\rho v \in \rho\,\mathsf{D}^q \cap \delta\,\Delta^{q+1}_{\Gamma_{\mathtt{N}}}$, e.g. $\delta\,\rho v = 0$, which we will not list here explicitly.

5 Inhomogeneous and More Boundary Conditions

In this section we will demonstrate that our error equalities also hold for Robin type boundary conditions, which means that our error equalities are true for many commonly used boundary conditions. Moreover, we emphasize that we can also handle inhomogeneous boundary conditions. Since it is clear that this method works in the general setting as well we will discuss it here just for the simple reaction-diffusion model problem from the introduction.

Let Ω be as in the latter section and now the boundary Γ be decomposed into three disjoint parts $\Gamma_{\mathtt{D}}$, $\Gamma_{\mathtt{N}}$ and $\Gamma_{\mathtt{R}}$. The model problem is: Find the scalar potential $u \in \mathsf{H}^1$ such that

$$\begin{aligned} -\operatorname{div} \nabla u + u &= f && \text{in } \Omega, \\ u &= g_1 && \text{on } \Gamma_{\mathtt{D}}, \\ n \cdot \nabla u &= g_2 && \text{on } \Gamma_{\mathtt{N}}, \\ n \cdot \nabla u + \gamma u &= g_3 && \text{on } \Gamma_{\mathtt{R}} \end{aligned}$$

hold. Hence, on $\Gamma_{\mathtt{D}}, \Gamma_{\mathtt{N}}$ and $\Gamma_{\mathtt{R}}$ we impose Dirichlet, Neumann and Robin type boundary conditions, respectively. In the Robin boundary condition, we assume that the coefficient $\gamma \geq \gamma_0 > 0$ belongs to L^∞. The dual variable for this problem is the flux $p := \nabla u \in \mathsf{D}$. Furthermore, as long as $\Gamma_{\mathtt{R}} \neq \emptyset$ and to avoid tricky discussions about traces and the corresponding $\mathsf{H}^{-1/2}$-spaces of Γ, $\Gamma_{\mathtt{D}}, \Gamma_{\mathtt{N}}$ and $\Gamma_{\mathtt{R}}$, which can be quite complicated, we assume for simplicity that $u \in \mathsf{H}^2$. Then, $p \in \mathsf{H}^1$ and all g_i belong to L^2 even to $\mathsf{H}^{1/2}$ of Γ. For the norms we simply have

$$\|(u, p)\|^2 = |u|^2_{\mathsf{H}^1} + |p|^2_{\mathsf{D}}.$$

Theorem 5.1. *For any approximation* $(\tilde{u}, \tilde{p}) \in \mathsf{H}^2 \times \mathsf{H}^1$ *with* $u - \tilde{u} \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}}$ *and* $p - \tilde{p} \in \mathsf{D}_{\Gamma_{\mathtt{N}}}$ *as well as* $n \cdot (p - \tilde{p}) + \gamma(u - \tilde{u}) = 0$ *on* $\Gamma_{\mathtt{R}}$

$$\|(u, p) - (\tilde{u}, \tilde{p})\|^2 + |u - \tilde{u}|^2_{\mathsf{L}^2(\Gamma_{\mathtt{R}}),\gamma} + |n \cdot (p - \tilde{p})|^2_{\mathsf{L}^2(\Gamma_{\mathtt{R}}),\gamma^{-1}} = \mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p})$$

holds with $\mathcal{M}_{\mathrm{mix}}$ *from Theorem 2.5. Moreover,* $|u - \tilde{u}|_{\mathsf{L}^2(\Gamma_{\mathtt{R}}),\gamma} = |n \cdot (p - \tilde{p})|_{\mathsf{L}^2(\Gamma_{\mathtt{R}}),\gamma^{-1}}$.

Proof. Following Remark 2.7 we have

$$\mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p}) = \underbrace{|u - \tilde{u}|^2_{\mathsf{H}^1} + |p - \tilde{p}|^2_{\mathsf{D}}}_{= \|(u,p) - (\tilde{u}, \tilde{p})\|^2} + 2\langle \nabla(u - \tilde{u}), \tilde{p} - p \rangle_{\mathsf{L}^2} + 2\langle u - \tilde{u}, \operatorname{div}(\tilde{p} - p) \rangle_{\mathsf{L}^2}.$$

Moreover, since $n \cdot (\tilde{p} - p)$ and $u - \tilde{u}$ belong to $\mathsf{L}^2(\Gamma)$ we have

$$\begin{aligned} &\langle \nabla(u - \tilde{u}), \tilde{p} - p \rangle_{\mathsf{L}^2} + \langle u - \tilde{u}, \operatorname{div}(\tilde{p} - p) \rangle_{\mathsf{L}^2} \\ &= \langle n \cdot (\tilde{p} - p), u - \tilde{u} \rangle_{\mathsf{L}^2(\Gamma)} = \langle n \cdot (\tilde{p} - p), u - \tilde{u} \rangle_{\mathsf{L}^2(\Gamma_{\mathtt{R}})} = \langle \gamma(u - \tilde{u}), u - \tilde{u} \rangle_{\mathsf{L}^2(\Gamma_{\mathtt{R}})}. \end{aligned}$$

As $\langle \gamma(u - \tilde{u}), u - \tilde{u} \rangle_{\mathsf{L}^2(\Gamma_{\mathtt{R}})} = \langle \gamma^{-1} n \cdot (p - \tilde{p}), n \cdot (p - \tilde{p}) \rangle_{\mathsf{L}^2(\Gamma_{\mathtt{R}})}$ we get the assertion. □

Remark 5.2. If all $g_i = 0$, we can set $(\tilde{u}, \tilde{p}) = (0,0)$ and get

$$\|(u,p)\|^2 + |u|^2_{\mathsf{L}^2(\Gamma_{\mathtt{R}}),\gamma} + |n \cdot p|^2_{\mathsf{L}^2(\Gamma_{\mathtt{R}}),\gamma^{-1}} = |f|^2_{\mathsf{L}^2},$$

which follows also directly from Remark 2.6 (ii'), $p = \nabla u$ and $n \cdot p = -\gamma u$ on $\Gamma_{\mathtt{R}}$ as well as

$$\begin{aligned} |f|^2_{\mathsf{L}^2} &= |\operatorname{div} p|^2_{\mathsf{L}^2} + |u|^2_{\mathsf{L}^2} - 2\langle \operatorname{div} \nabla u, u\rangle_{\mathsf{L}^2} \\ &= |\operatorname{div} p|^2_{\mathsf{L}^2} + |u|^2_{\mathsf{L}^2} + 2|\nabla u|_{\mathsf{L}^2} - 2\langle n \cdot \nabla u, u\rangle_{\mathsf{L}^2(\Gamma)} \\ &= |\operatorname{div} p|^2_{\mathsf{L}^2} + |u|^2_{\mathsf{L}^2} + 2|\nabla u|_{\mathsf{L}^2} - 2\underbrace{\langle n \cdot \nabla u, u\rangle_{\mathsf{L}^2(\Gamma_{\mathtt{R}})}}_{= -|u|^2_{\mathsf{L}^2(\Gamma_{\mathtt{R}}),\gamma}}. \end{aligned}$$

Thus, in this case the assertion of Theorem 5.1 has a normalized counterpart as well.

If $\Gamma_{\mathtt{R}} = \emptyset$ we have a pure mixed Dirichlet and Neumann boundary.

Theorem 5.3. *Let $\Gamma_{\mathtt{R}} = \emptyset$. For any approximation $(\tilde{u}, \tilde{p}) \in \mathsf{H}^1 \times \mathsf{D}$ with $u - \tilde{u} \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}}$ and $p - \tilde{p} \in \mathsf{D}_{\Gamma_{\mathtt{N}}}$*

$$\|(u,p) - (\tilde{u}, \tilde{p})\|^2 = \mathcal{M}_{\mathrm{mix}}(\tilde{u}, \tilde{p})$$

holds with $\mathcal{M}_{\mathrm{mix}}$ from Theorem 2.5.

Corollary 5.4. *Let $\Gamma_{\mathtt{R}} = \emptyset$. Theorem 5.3 provides the well known a posteriori error estimates for the primal and dual problems.*

(i) *For any $\tilde{u} \in \mathsf{H}^1$ with $u - \tilde{u} \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}}$ it holds $|u - \tilde{u}|^2_{\mathsf{H}^1} = \min\limits_{\substack{\psi \in \mathsf{D} \\ p - \psi \in \mathsf{D}_{\Gamma_{\mathtt{N}}}}} \mathcal{M}_{\mathrm{mix}}(\tilde{u}, \psi) = \mathcal{M}_{\mathrm{mix}}(\tilde{u}, p)$.*

(ii) *For any $\tilde{p} \in \mathsf{D}$ with $p - \tilde{p} \in \mathsf{D}_{\Gamma_{\mathtt{N}}}$ it holds $|p - \tilde{p}|^2_{\mathsf{D}} = \min\limits_{\substack{\varphi \in \mathsf{H}^1 \\ u - \varphi \in \mathsf{H}^1_{\Gamma_{\mathtt{D}}}}} \mathcal{M}_{\mathrm{mix}}(\varphi, \tilde{p}) = \mathcal{M}_{\mathrm{mix}}(u, \tilde{p})$.*

6 Numerical Examples

In this section we show by some academic test cases the numerical performance of our error equalities. All the calculations have been done using MATLAB, and the reported values in the tables have not been rounded, but are simply cut-offs of values reported by MATLAB. The main quantity of interest is the difference between the exact error and the value given by the majorant for a certain approximation $(\tilde{u}, \tilde{p})$, i.e.,

$$\delta := \big| \|(u,p) - (\tilde{u}, \tilde{p})\| - \mathcal{M}_{\dots}(\tilde{u}, \tilde{p})^{1/2} \big|,$$

where the test problems are either from the reaction-diffusion problems from Section 4.1 or from the eddy-current problems from Sections 4.2 and 4.3. Where the finite element method (FEM) has been used, we have employed only linear triangular elements in 2D and linear tetrahedral elements in 3D. In all the examples below we calculated the approximations $\tilde{u}$ and $\tilde{p}$ (or $\tilde{E}$ and $\tilde{H}$) in the same mesh only for the sake of convenience. Using different meshes for the primal and dual approximations is allowed. We also used only regular meshes, but irregular meshes can be used as well. The only requirement is that the approximations must be conforming, meaning that they belong to the appropriate Sobolev spaces and fulfill the boundary conditions exactly. All finite element solvers were implemented in the vectorized manner explained in [13].

Example 6.1. We take the 3D-reaction-diffusion problem from Section 4.1 and choose the unit cube $\Omega := (0,1)^3$ with exact solution

$$u(x) := \prod_{i=1}^{3} x_i(1-x_i),$$

where u satisfies the zero Dirichlet boundary conditions on the whole boundary, i.e., $\Gamma_{\mathtt{D}} = \Gamma$ and $\Gamma_{\mathtt{N}} = \emptyset$, and the following data

$$\alpha(x) := \alpha := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 10 \end{bmatrix}, \quad \rho(x) := \begin{cases} 1 & \text{if} \quad 0 < x_1 < 1/4 \\ 10 & \text{if} \quad 1/4 < x_1 < 3/4 \\ 25 & \text{if} \quad 3/4 < x_1 < 1 \end{cases}.$$

This means that the approximation of the dual variable does not have any boundary condition. We calculated the approximation globally by solving the primal and dual problem with standard linear Courant elements and linear Raviart-Thomas elements, respectively. We will denote this finite element approximation pair by $(u_{\mathtt{h}}, p_{\mathtt{h}})$. The resulting linear systems were solved directly in MATLAB. The approximations were calculated in uniformly refined regular meshes, where the jumps in the reaction coefficient ρ coincide with element boundaries. For each mesh we computed the exact combined error and the majorant $\mathcal{M}_{\mathrm{rd}}(u_{\mathtt{h}}, p_{\mathtt{h}})$. The results are displayed in Table 6.1. The first column shows the number of elements $N_{\mathtt{elem}}$ of the mesh. The second and third column show the exact error and the value given by the majorant. The fourth column shows the difference δ between the exact error and the value given by the majorant.

Table 6.1: Example 6.1 (3D-reaction-diffusion)

$N_{\mathtt{elem}}$	$\|(u,p)-(u_{\mathtt{h}},p_{\mathtt{h}})\|$	$\mathcal{M}_{\mathrm{rd}}(u_{\mathtt{h}},p_{\mathtt{h}})^{1/2}$	difference δ
384	0.12803218100	0.12803218100	5.551115123e-17
3072	0.06736516349	0.06736516349	4.163336342e-17
24576	0.03433600867	0.03433600867	9.714451465e-17
196608	0.01728806289	0.01728806289	3.469446952e-18

Example 6.2. This test is similar to the Example 1 except that the linear systems resulting from the finite element computations were not solved directly, but with an iterative method, where the stopping tolerance was set to the crude value of 10^{-4}. The approximation pair obtained by this method is denoted by $(u_{\mathtt{iter}}, p_{\mathtt{iter}})$. No preconditioning was done. The iterative solver of the linear system of the dual problem converged only for the smallest mesh, and the error actually grows between the two last meshes. With this stopping tolerance this is expected and was purposefully done so in order to obtain approximations which are relatively far from having the Galerkin orthogonality property. We did this test simply to demonstrate that Galerkin orthogonality is not a requirement for the equality to hold. The results are displayed in Table 6.2.

Table 6.2: Example 6.2 (3D-reaction-diffusion)

$N_{\mathtt{elem}}$	$\|(u,p) - (u_{\mathtt{iter}}, p_{\mathtt{iter}})\|$	$\mathcal{M}_{\mathrm{rd}}(u_{\mathtt{iter}}, p_{\mathtt{iter}})^{1/2}$	difference δ
384	0.12803483290	0.12803483290	2.775557562e-17
3072	0.06868358511	0.06868358511	6.938893904e-17
24576	0.05294561599	0.05294561599	6.245004514e-17
196608	0.09166231565	0.09166231565	9.714451465e-17

Example 6.3. We ran the problem data of Example 6.1 with subsequently refined regular meshes, where the approximation of the primal variable $u_{\mathtt{h}}$ was again obtained by the linear Courant finite elements. The resulting linear system was solved directly. The approximation of the dual variable was calculated by averaging the values $\alpha \nabla u_{\mathtt{h}}$ to the nodes of the mesh. This procedure is often called the gradient averaging method and we will denote the resulting function by $p_{\mathtt{avg}}$. The results can be seen in Table 6.3.

Table 6.3: Example 6.3 (3D-reaction-diffusion)

$N_{\mathtt{elem}}$	$\|(u,p) - (u_{\mathtt{h}}, p_{\mathtt{avg}})\|$	$\mathcal{M}_{\mathrm{rd}}(u_{\mathtt{h}}, p_{\mathtt{avg}})^{1/2}$	difference δ
384	0.2698605861	0.2698605861	0
3072	0.2285323585	0.2285323585	0
24576	0.1831121412	0.1831121412	6.106226635e-16
196608	0.1333268308	0.1333268308	1.693090113e-15

Example 6.4. We take the 2D-eddy-current problem from Section 4.3 and choose the unit square $\Omega := (0,1)^2$ with $\epsilon = \mathrm{id}$ and $\mu = 1$. We split the domain in the two parts $\Omega_1 := \{x \in \Omega \mid x_1 > x_2\}$ and $\Omega_2 = \Omega \setminus \overline{\Omega_1}$ in order to define the following discontinuous solution

$$E|_{\Omega_1}(x) := \begin{bmatrix} \sin(2\pi x_1) + 2\pi\cos(2\pi x_1)(x_1 - x_2) \\ \sin\big((x_1 - x_2)^2(x_1 - 1)^2 x_2\big) - \sin(2\pi x_1) \end{bmatrix}, \quad E|_{\Omega_2}(x) := 0.$$

Note that indeed $E \in \mathsf{R} \setminus \mathsf{H}^1$ and $\operatorname{rot} E \in \mathsf{H}^1$ with

$$\operatorname{rot} E|_{\Omega_1}(x) = 2x_2(x_1 - x_2)(x_1 - 1)(2x_1 - x_2 - 1)\cos(2\pi x_1).$$

We set zero Neumann boundary conditions on the whole boundary, i.e., $\Gamma_{\mathtt{D}} = \emptyset$ and $\Gamma_{\mathtt{N}} = \Gamma$. The exact solution and its rotation is visualized in Figure 6.1. We calculated the approximation globally by solving the primal and dual problem with linear Nédélec elements and linear Courant elements, respectively. This finite element approximation pair will be denoted by $(E_{\mathtt{h}}, H_{\mathtt{h}})$. The resulting linear systems were solved directly. The approximations were calculated in uniformly refined regular meshes, where the jumps in the exact solution and in the right hand side J coincide with element boundaries. For each mesh we calculated the exact combined error and the majorant $\mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})$. The results are displayed in Table 6.4.

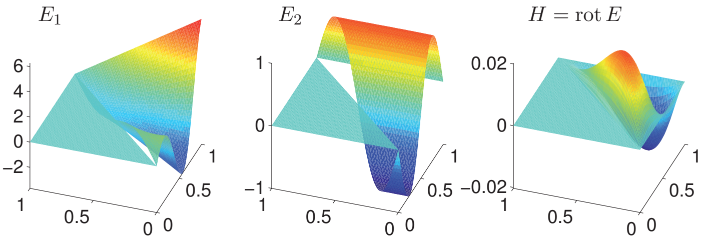


Figure 6.1: The two components of the exact solution E and its rotation H of Example 6.4.

Table 6.4: Example 6.4 (2D-eddy-current)

$N_{\mathtt{elem}}$	$\|(E,H) - (E_{\mathtt{h}}, H_{\mathtt{h}})\|$	$\mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})^{1/2}$	difference δ
800	0.151485078300	0.151485078300	2.220446049e-16
3200	0.075877018950	0.075877018950	0
12800	0.037956449900	0.037956449900	7.632783294e-17
51200	0.018980590110	0.018980590110	6.938893904e-17
204800	0.009490605462	0.009490605462	2.602085214e-17

Example 6.5. We take the 3D-eddy-current problem from Section 4.2 and choose the unit cube $\Omega := (0,1)^3$ with $\epsilon = \mu = \mathrm{id}$. Again we split the domain in the two parts $\Omega_1 := \{x \in \Omega \mid x_1 > x_2\}$ and $\Omega_2 = \Omega \setminus \overline{\Omega_1}$ in order to define the following discontinuous solution

$$E(x) := \chi_{\Omega_1}(x) \begin{bmatrix} \sin(2\pi x_1) + 2\pi \cos(2\pi x_1)(x_1 - x_2) \\ \sin\big((x_1 - x_2)^2(x_1-1)^2 x_2\big) - \sin(2\pi x_1) \\ 0 \end{bmatrix} + \xi(x) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix},$$

where $\xi(x) := \prod_{i=1}^3 x_i^2(1-x_i)^2$. Thus, we extended the discontinuous vector field of Example 6.4 by zero in the third component and added a smooth bubble in the third component. Hence, $E \in \mathsf{R} \setminus \mathsf{H}^1$ and $\operatorname{rot} E \in \mathsf{R}$ with

$$\operatorname{rot} E(x) = \chi_{\Omega_1}(x)\big(2x_2(x_1-x_2)(x_1-1)(2x_1-x_2-1)\cos(2\pi x_1)\big) \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} + \begin{bmatrix} \partial_2 \xi \\ -\partial_1 \xi \\ 0 \end{bmatrix}(x).$$

Note that even $\operatorname{rot} E \in \mathsf{H}^1$ holds. We set zero Neumann boundary conditions on the whole boundary, i.e., $\Gamma_{\mathtt{D}} = \emptyset$ and $\Gamma_{\mathtt{N}} = \Gamma$. We calculated the approximation globally by solving the primal and dual problem with linear Nédélec elements. This finite element approximation pair will be denoted by $(E_{\mathtt{h}}, H_{\mathtt{h}})$. The resulting linear systems were solved directly. The approximations were calculated in uniformly refined regular meshes, where the jumps in the exact solution and in the right hand side J coincide with element boundaries. For each mesh we calculated the exact combined error and the majorant $\mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})$. The results are displayed in Table 6.5.

Table 6.5: Example 6.5 (3D-eddy-current)

$N_{\mathtt{elem}}$	$\\|(E,H) - (E_{\mathtt{h}}, H_{\mathtt{h}})\\|$	$\mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})^{1/2}$	difference δ
384	0.7228185218	0.7228185218	3.330669074e-16
3072	0.3717887807	0.3717887807	6.106226635e-16
24576	0.1883612515	0.1883612515	2.775557562e-16
196608	0.0945757836	0.0945757836	8.604228441e-16

Example 6.6. We take the problem data of Example 6.4 and solve the primal and dual problems in adaptively refined meshes with linear Nédélec elements and linear Courant elements, respectively. This finite element approximation pair will be denoted by $(E_{\mathtt{h}}, H_{\mathtt{h}})$ and the linear systems are solved directly. We compare optimal refinement achieved by using the exact error distribution e_T to the refinement provided by the distribution of the majorant η_T, where

$$\begin{aligned} e_T^2 &:= \|(E,H) - (E_{\mathtt{h}}, H_{\mathtt{h}})\|_T^2 := |E - E_{\mathtt{h}}|^2_{\mathsf{R}(T)} + |H - H_{\mathtt{h}}|^2_{\mathsf{H}^1(T)}, \\ \eta_T^2 &:= \mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})_T := |J - E_{\mathtt{h}} - \nabla^{\perp} H_{\mathtt{h}}|^2_{\mathsf{L}^2(T)} + |H_{\mathtt{h}} - \operatorname{rot} E_{\mathtt{h}}|^2_{\mathsf{L}^2(T)} \end{aligned}$$

and T denotes an element (triangle) of the mesh discretization. We start from a regular mesh with 200 elements, and perform nine refinement iterations, where on each iteration 30% of elements with the highest amount of error are refined. The refinement of element meshes is done by regular refinement such that the resulting mesh does not contain hanging nodes. The results of Figure 6.2 show that even though the equality is *global*, the majorant can still be used to perform reliable adaptive computations. We see from Table 6.6 that the number of elements in the optimal meshes and the meshes produced using η_T are very close to each other. In Figure 6.3 we have depicted the meshes after the fourth refinement. Figure 6.4 depicts one of the finest parts of the final meshes. In fact, the adaptive refinement using η_T is very close to optimal in each step, and the resulting approximation after the last refinement is practically the same.

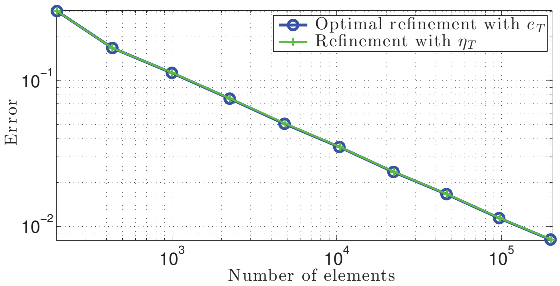


Figure 6.2: Adaptive computation of Example 6.6, where the error is measured in the combined norm.

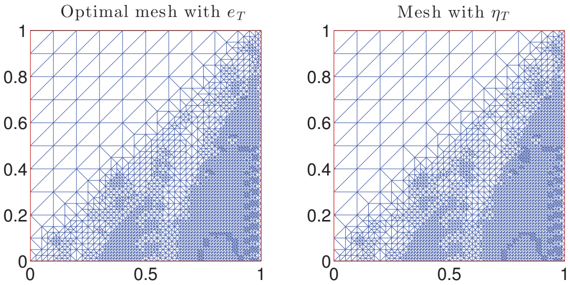


Figure 6.3: Adaptive mesh after the fourth refinement in Example 6.6. There are 4823 elements in the optimal mesh, and 4878 elements in the mesh calculated with the help of η_T.

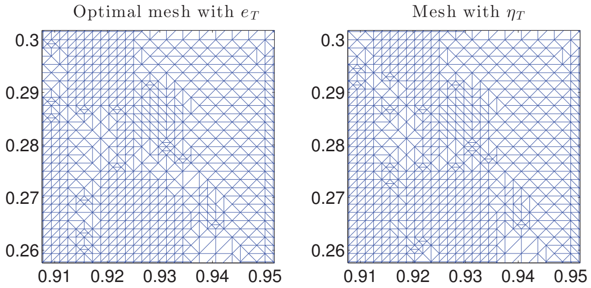


Figure 6.4: One of the most fine parts in the final adaptive mesh in Example 6.6.

Table 6.6: Adaptive computation of Example 6.6. The number of elements in the optimal meshes and the meshes generated by the help of η_T.

Ref.	optimal	with η_T	difference	difference %
-	200	200	0	0
1	434	434	0	0
2	998	1002	4	0.40
3	2240	2252	12	0.53
4	4823	4878	55	1.14
5	10378	10446	68	0.65
6	22116	22337	221	0.99
7	46388	46768	380	0.81
8	96859	97832	973	1.00
9	198704	200970	2266	1.14

Example 6.7. We take the 2D-eddy-current problem of Section 4.3 in the L-shaped domain $\Omega := (0,1)^2 \setminus \big([1/2, 1] \times [0, 1/2]\big)$ with $\epsilon = \mathrm{id}$, $\mu = 1000$ and $J = [1,0]^\top$. We set zero Dirichlet boundary conditions on the whole boundary, i.e., $\Gamma_{\mathtt{D}} = \Gamma$ and $\Gamma_{\mathtt{N}} = \emptyset$. The exact solution of this problem is unknown. However, since the majorant gives indeed the exact error in the combined norm, we will use this information in this example. Therefore, all the error values in Figure 6.5 and Table 6.7 are values of the majorant. We compare uniform refinement and adaptive refinement using η_T with

$$\eta_T^2 = \mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})_T = |J - E_{\mathtt{h}} - \nabla^\perp H_{\mathtt{h}}|^2_{\mathtt{L}^2(T)} + |H_{\mathtt{h}} - \mu^{-1} \operatorname{rot} E_{\mathtt{h}}|^2_{\mathtt{L}^2(T),\mu},$$

refining 30% of elements on each refinement iteration as before. We solve the primal and dual problems with linear Nédélec elements and linear Courant elements, respectively. The resulting linear systems are solved directly. We see from Figure 6.5 that the adaptive procedure is beneficial in this example. We have also depicted the approximation in Figure 6.6 and the mesh in Figure 6.7 after the fifth refinement.

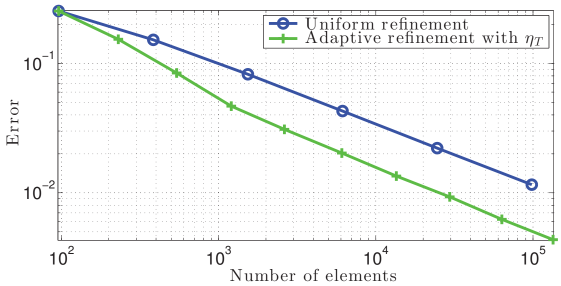


Figure 6.5: Adaptive computation of Example 6.7.

Table 6.7: Example 6.7 (2D-eddy-current) Adaptively refined meshes.

$N_{\mathtt{elem}}$	$\mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})^{1/2}$	$\mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})^{1/2}/\|J\|_{\mathtt{L}^2}$
96	0.2534	0.2926
230	0.1534	0.1771
541	0.0842	0.0973
1204	0.0467	0.0539
2623	0.0309	0.0357
6082	0.0203	0.0234
13514	0.0135	0.0155
29530	0.0093	0.0107
63363	0.0062	0.0072
134205	0.0043	0.0050

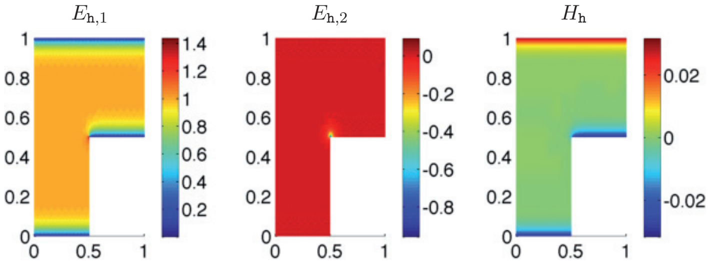


Figure 6.6: The two components of the approximate primal variable E_{h} and the dual variable H_{h} of Example 6.7 after the third adaptive refinement.

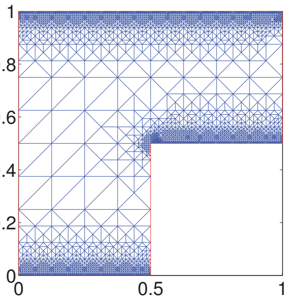


Figure 6.7: Adaptive mesh after the fifth adaptive refinement in Example 6.7.

Example 6.8. We take the 2D-eddy-current problem of Section 4.3 in $\Omega := (0,1)^2$. In order to define discontinuous data, we define with $\xi(x) := \ln(2 + x_2)$ and

$$\begin{aligned} &\Omega_1 := \big((0,1)\times(0.4,0.6)\big)\cup\big((0.3,0.5)\times(0,1)\big), && \epsilon|_{\Omega_1} := \mathrm{id}, && \epsilon|_{\Omega\setminus\overline{\Omega}_1} := 100\cdot\mathrm{id},\\ & && \mu|_{\Omega_1} := 1000, && \mu|_{\Omega\setminus\overline{\Omega}_1} := 1,\\ &\Omega_2 := (0,1)\times(0.35,0.65), && J|_{\Omega_2} := \xi\begin{bmatrix}1\\0\end{bmatrix}, && J|_{\Omega\setminus\overline{\Omega}_2} := -\xi\begin{bmatrix}0\\1\end{bmatrix}.\end{aligned}$$

We set zero Dirichlet boundary conditions on the right side of the boundary and zero Neumann boundary condition on the remaining part, i.e., $\Gamma_{\mathtt{D}} = \{x \in \Omega \mid x_1 = 1\}$. As in Example 6.7, the exact solution of this problem is unknown, so the error values in Figure 6.8 and Table 6.8 are the values of the majorant. We compare uniform refinement and adaptive refinement using η_T with

$$\eta_T^2 = \mathcal{M}_{\mathrm{ec}}(E_{\mathtt{h}}, H_{\mathtt{h}})_T = |J - \epsilon E_{\mathtt{h}} - \nabla^{\perp} H_{\mathtt{h}}|^2_{\mathtt{L}^2(T),\epsilon^{-1}} + |H_{\mathtt{h}} - \mu^{-1}\operatorname{rot} E_{\mathtt{h}}|^2_{\mathtt{L}^2(T),\mu},$$

refining 30% of elements on each refinement iteration as before. We solve the primal and dual problems with linear Nédélec elements and linear Courant elements, respectively. The resulting linear systems are solved directly. Again, we see from Figure 6.8 that the adaptive procedure is beneficial in this example. We have also depicted the approximation in Figure 6.9 and the mesh in Figure 6.10 after the third refinement.

To conclude, in all the tests performed, nonzero values of δ were of magnitude 10^{-18}-10^{-15}. This is within the limit of machine precision, so numerically these numbers are considered zero. In addition to verifying the equality, we also performed three simple examples to show that the majorant can be used to perform refinement of element meshes without any additional computational expenditures.

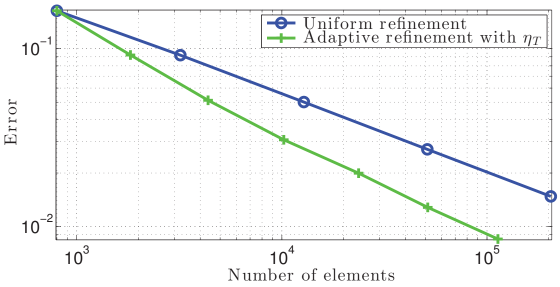


Figure 6.8: Adaptive computation of Example 6.8.

Table 6.8: Example 6.8 (2D-eddy-current) Adaptively refined meshes.

$N_{\mathtt{elem}}$	$\mathcal{M}_{\mathrm{ec}}(E_{\mathrm{h}}, H_{\mathrm{h}})^{1/2}$	$\mathcal{M}_{\mathrm{ec}}(E_{\mathrm{h}}, H_{\mathrm{h}})^{1/2}/\lvert J\rvert_{\mathsf{L}^2,\epsilon^{-1}}$
800	0.1632	0.2941
1827	0.0921	0.1659
4367	0.0513	0.0924
10214	0.0307	0.0554
23657	0.0199	0.0359
51429	0.0128	0.0231
113073	0.0085	0.0153

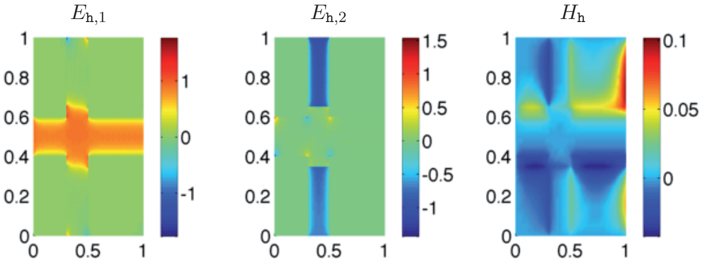


Figure 6.9: The two components of the approximate primal variable E_{h} and the dual variable H_{h} of Example 6.8 after the third adaptive refinement.

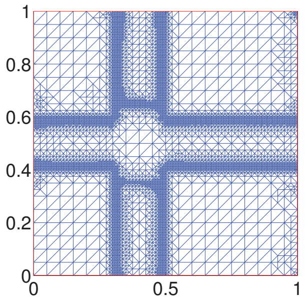


Figure 6.10: Adaptive mesh after the third adaptive refinement in Example 6.8.

Acknowledgements We are deeply indebted to Sergey Repin for so many interesting and encouraging discussions. We thank him gratefully not only for being an academic colleague and teacher but also a good friend.

The first author also thanks Jan Valdman for showing how to program vectorized FEM solvers. The first author is funded by the Finnish foundations KAUTE Foundation and Väisälä Foundation of the Finnish Academy of Science and Letters.

This contribution has been worked out mainly while the first author was visiting the Fakultät für Mathematik of the Universität Duisburg-Essen during 2013.

References

[1] I. Anjam, O. Mali, A. Muzalevskiy, P. Neittaanmäki, and S. Repin. A posteriori error estimates for a Maxwell type problem. *Russian J. Numer. Anal. Math. Modelling*, 24(5):395–408, 2009.

[2] F. Brezzi and M. Fortin. *Mixed and hybrid finite element methods*. Springer, New York, 1991.

[3] V. Gol'dshtein, I. Mitrea, and M. Mitrea. Hodge decompositions with mixed boundary conditions and applications to partial differential equations on Lipschitz manifolds. *J. Math. Sci. (N.Y.)*, 172(3):347–400, 2011.

[4] T. Jakab, I. Mitrea, and M. Mitrea. On the regularity of differential forms satisfying mixed boundary conditions in a class of Lipschitz domains. *Indiana Univ. Math. J.*, 58(5):2043–2071, 2009.

[5] F. Jochmann. A compactness result for vector fields with divergence and curl in $L^q(\Omega)$ involving mixed boundary conditions. *Appl. Anal.*, 66:189–203, 1997.

[6] P. Kuhn. *Die Maxwellgleichung mit wechselnden Randbedingungen*. Dissertation, Universität Essen, Fachbereich Mathematik, http://arxiv.org/abs/1108.2028, *Shaker*, 1999.

[7] O. Mali, A. Muzalevskiy, and D. Pauly. Conforming and non-conforming functional a posteriori error estimates for elliptic boundary value problems in exterior domains: Theory and numerical tests. *Russian J. Numer. Anal. Math. Modelling*, 28(6):577–596, 2013.

[8] O. Mali, P. Neittaanmäki, and S. Repin. *Accuracy verification methods, theory and algorithms*. Springer, 2014.

[9] P. Neittaanmäki and S. Repin. *Reliable methods for computer simulation, error control and a posteriori estimates*. Elsevier, New York, 2004.

[10] D. Pauly and S. Repin. Functional a posteriori error estimates for elliptic problems in exterior domains. *J. Math. Sci. (N.Y.)*, 162(3):393–406, 2009.

[11] D. Pauly and S. Repin. Two-sided a posteriori error bounds for electro-magneto static problems. *J. Math. Sci. (N.Y.)*, 166(1):53–62, 2010.

[12] D. Pauly, S. Repin, and Rossi T. Estimates for deviations from exact solutions of Maxwell's initial boundary value problem. *Ann. Acad. Sci. Fenn. Math.*, 36(2):661–676, 2011.

[13] T. Rahman and J. Valdman. Fast MATLAB assembly of FEM matrices in 2D and 3D: nodal elements. *Appl. Math. Comput.*, 219(13):7151–7158, 2013.

[14] S. Repin. *A posteriori estimates for partial differential equations*. Walter de Gruyter (Radon Series Comp. Appl. Math.), Berlin, 2008.

[15] S. Repin. Estimates of deviations from exact solutions of initial boundary value problems for the wave equation. *J. Math. Sci. (N. Y.)*, 159(2):229–240, 2009.

[16] S. Repin, S. Sauter, and A. Smolianski. Two-sided a posteriori error estimates for mixed formulations of elliptic problems. *SIAM J. Numer. Anal.*, 45(3):928–945, 2007.

PIV

NEW INDICATORS OF APPROXIMATION ERRORS FOR PROBLEMS IN CONTINUUM MECHANICS

by

I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin (2010)

J. C. F. Pereira, A. Sequeira and J. M. C. Pereira (Eds.), Proceedings of the V
European Conference on Computational Fluid Dynamics ECCOMAS CFD 2010,
Lisbon, Portugal,14-17 June 2010, ISBN: 978-989-96778-1-4

V European Conference on Computational Fluid Dynamics
ECCOMAS CFD 2010
J. C. F. Pereira and A. Sequeira (Eds)
Lisbon, Portugal,14-17 June 2010

NEW INDICATORS OF APPROXIMATION ERRORS FOR PROBLEMS IN CONTINUUM MECHANICS

Immanuel B. Anjam[1], **Olli J. Mali**[2],
Pekka J. Neittaanmäki[3] **and Sergey I. Repin**[4]

[1,2,3]University of Jyväskylä,
Department of Mathematical Information Technology,
P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Finland
e-mail: {immanuel.anjam,olli.mali,pn}@jyu.fi

[4]V. A. Steklov Institute of Mathematics in St. Petersburg
Fontanka 27, RU-191024, St. Petersburg, Russia
e-mail: repin@pdmi.ras.ru

Key words: continuum mechanics, a posteriori, error indicator, reliable numerical methods

Abstract. *In this paper we present a new error indicator for approximate solutions of elliptic problems. We discuss error indication with the paradigm of the diffusion problem, however the techniques are easily adaptable to more complicated elliptic problems, for example to linear elasticity, viscous flow models and electromagnetic models. The proposed indicator does not contain mesh dependent constants and it admits parallelization.*

1 INTRODUCTION

Various a posteriori indicators of approximation errors are widely used in computer simulation. Error indicators for finite element approximations are usually based either on evaluation of a weak residual norm or on post-processing (e.g., gradient averaging) and are applicable only to Galerkin approximations. In this paper, we discuss a different class of error indicators that follow from a posteriori estimates of the functional type (a consequent exposition of the corresponding theory is given in the books[5,7]). These estimates do not contain mesh dependent constants, do not exploit specific properties of the numerical method or approximations used and are valid for any conforming approximation.

In this paper, we modify some ideas of the functional approach and derive error indicators of a new type. The indicators contain the corresponding numerical solution v, problem data and an arbitary function y, which is to be selected in a suitable way. For this task, we apply two different methods (global and local) and compare their efficiency.

Let Ω be a bounded and connected domain in $\mathbb{R}^d$ with Lipschitz boundary $\partial\Omega$. Consider the the following problem: find a scalar function u such that

$$-\mathrm{div}\,\mathbf{A}\nabla u = f \qquad \text{in } \Omega, \tag{1}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{2}$$

where $\mathbf{A}$ is a symmetric $d \times d$ matrix with coefficients in $L^\infty(\Omega)$ and $f \in L^2(\Omega)$. The generalized solution to this problem is a function $u \in \overset{\circ}{H}{}^1(\Omega)$ that satisfies the relation

$$\int_\Omega \mathbf{A}\nabla u \cdot \nabla w \, dx = \int_\Omega f w \, dx, \qquad \forall w \in \overset{\circ}{H}{}^1(\Omega), \tag{3}$$

where $\overset{\circ}{H}{}^1(\Omega)$ is the space of functions from $H^1(\Omega)$ which vanish on $\partial\Omega$. For this problem the natural energy norm is defined as

$$|\!|\!| \, u \, |\!|\!|^2 := \|\nabla u\|^2_{\mathbf{A}} := \int_\Omega \mathbf{A}\nabla u \cdot \nabla u \, dx.$$

We denote by $\|\cdot\|$ the L_2 norm of scalar- and vector-valued functions.

2 ERROR MAJORANT AND INDICATOR

Guaranteed error bounds for the problem (1)-(2) are derived by transformations of the integral identity (3), which lead to the following result[5,7].

Proposition 2.1. *Let u be the exact solution and $v \in \overset{\circ}{H}{}^1(\Omega)$ a numerical solution to the problem* (1)-(2)*. Then*

$$|\!|\!| \, u - v \, |\!|\!| \leq M_\oplus(v, y)\,, \qquad \forall y \in H(\mathrm{div}, \Omega),$$

where

$$M_\oplus(v, y) := C_\Omega \|f + \operatorname{div} y\| + \|y - \mathbf{A}\nabla v\|_{\mathbf{A}^{-1}}. \tag{4}$$

Here C_Ω is the constant in the Friedrichs inequality and $y \in H(\operatorname{div}, \Omega)$ is an arbitary function.

The quality of the majorant (4) depends on how well the arbitary function y represents the exact flux $p = \mathbf{A}\nabla u$. This estimate does not contain a gap between the exact error and the estimate. This fact is easy to establish by replacing y with the exact flux. The first term vanishes and the majorant becomes

$$M_\oplus(v, p) = \|\nabla(u - v)\|_{\mathbf{A}} = \| u - v \| .$$

Indeed, if the free function y is chosen properly, the first term of majorant (4) is small. Therefore it is reasonable to assume that we can define the following error indicator from the latter term of the majorant.

Proposition 2.2. *Let u be the exact solution and $v \in \overset{\circ}{H}{}^1(\Omega)$ a numerical solution to the problem* (1)-(2)*. We define the error indicator*

$$I(v, y) := \|y - \mathbf{A}\nabla v\|^2, \tag{5}$$

where $y \in H(\operatorname{div}, \Omega)$ is an arbitary function. The indicator I estimates the distribution of $\| u - v \|^2$ in the domain Ω.

3 OBTAINING THE ARBITARY FUNCTION y

The majorant and the indicator contain the arbitary function y, which we call the flux. In this section we show several ways how to obtain this parameter for the diffusion problem. These same methods can be used also for other elliptic problems.

The problem of finding y for the diffusion problem burns down to approximating the exact flux $p = \mathbf{A}\nabla u$. There are several ways to obtain estimates to the exact flux. First we discuss the global minimization technique, and then we propose a (new) local minimization procedure.

3.1 Global minimization

One way to obtain good approximations for the exact flux is to minimize the majorant $M_\oplus$ defined by (4) globally with respect to y. For this we transform the majorant to a quadratic form. This is done by squaring the majorant and using the algebraic inequality $(a+b)^2 \leq (1+\beta)a^2 + (1+\frac{1}{\beta})b^2$ which holds for all $\beta > 0$. The estimate proposed in 2.1 becomes

$$\| u - v \|^2 \leq \mathcal{M}(v, y, \beta) := (1+\beta)C_\Omega^2\|f + \operatorname{div} y\|^2 + \left(1 + \frac{1}{\beta}\right)\|y - \mathbf{A}\nabla v\|^2_{\mathbf{A}^{-1}}. \tag{6}$$

Minimizing (6) globally results in the following finite element problem for $y \in H(\mathrm{div}, \Omega)$:

$$
\begin{aligned}
(1+\beta)C_\Omega^2 \int_\Omega \mathrm{div}\, y \,\mathrm{div}\, \phi \, dx + \left(1 + \frac{1}{\beta}\right) \int_\Omega \mathbf{A}^{-1} y \cdot \phi \, dx = \\
= -(1+\beta)C_\Omega^2 \int_\Omega f \mathrm{div}\, \phi \, dx + \left(1 + \frac{1}{\beta}\right) \int_\Omega \phi \cdot \nabla v \, dx. \quad \forall \phi \in H(\mathrm{div}, \Omega).
\end{aligned}
$$

A natural choise to solve this problem is to use Raviart-Thomas elements[4,6]. This method produces good approximations for the exact flux, but is relatively time-consuming. For error indication purposes less expensive methods are preferrable.

3.2 Averaging procedures

A very popular method to approximate the exact flux is to post-process the approximate flux $\mathbf{A}\nabla v$[1,3,8]. If v belongs to the space $H^1(\Omega)$, then its gradient ∇v is constant in each element. If also the matrix $\mathbf{A}$ is constant in each element, we can apply very simple averaging procedures to the approximate flux.

A common way is to average the approximate flux to nodes: for each node, calculate $\mathbf{A}\nabla v$ in each related element and average the values weighted by the areas of respective elements. We denote this procedure by G_N.

It is also possible to average the normal components of the approximate flux. In 2D these values are averaged to edges of elements. Let c_{nl} denote the unknown degree of freedom related to edge e_{nl} with edge length $|e_{nl}|$. Here the subindex letters n and l denote the numbers of the nodes which define the edge. We denote by T_{knl}, T_{nml} the elements related to this edge and by n_{knl}, n_{nml} their respective unit outward normals on the boundary. This setting is visualized in Figure 1. The following equation averages the normal component of $\mathbf{A}\nabla v$ to the edge e_{nl}:

$$
c_{nl} = \frac{|e_{nl}| \left(\mathbf{A}\nabla v|_{T_{knl}} \cdot n_{knl} - \mathbf{A}\nabla v|_{T_{nml}} \cdot n_{nml} \right)}{2} \tag{7}
$$

In 3D the normal components are averaged in a similiar way. The only difference is that now we average the values to faces instead of edges. Let c_{nlm} denote the unknown degree of freedom related to face f_{nlm} whose area is $|f_{nlm}|$. We denote by T_{knlm}, T_{omln} the elements related to this face and by n_{knlm}, n_{omln} their respective unit outward normals on the boundary, see Figure 1. The following equation averages the normal component of $\mathbf{A}\nabla v$ to the face f_{nlm}:

$$
c_{nlm} = \frac{|f_{nlm}| \left(\mathbf{A}\nabla v|_{T_{knlm}} \cdot n_{knlm} - \mathbf{A}\nabla v|_{T_{omln}} \cdot n_{omln} \right)}{2} \tag{8}
$$

We denote by G_{RT} the procedure, which calculates the values of (7) for all edges or (8) for all faces in a given mesh. It should be noted that the operator G_{RT} essentially produces functions from linear Raviart-Thomas finite element space.

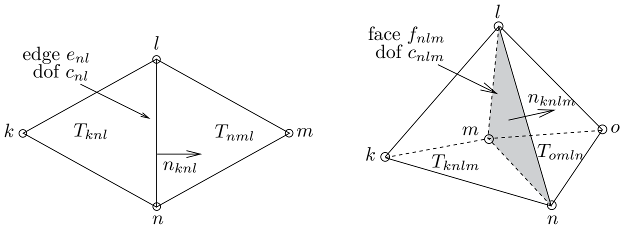


Figure 1: Two neighboring elements in 2D and 3D.

3.3 A post-processing method

In this section we present a post-processing method which gives even better approximations for the exact flux $\mathbf{A}\nabla u$. Assume that the initial approximation is obained by the averaging operator G_{RT} defined in the previous section. In our previous paper[2] we made further post-processing of the approximate flux (for the poisson problem) by minimizing only a part of the majorant. In this paper we take this same idea further and choose to post-process $y = G_{RT}(\mathbf{A}\nabla v)$ by minimizing the whole majorant $\mathcal{M}_\oplus$ on every pair of neghboring triangular elements. Next we show how to do this in 3D (the 2D case is very similar).

Since y is computed by the averaging operator G_{RT}, it can be represented as

$$y = \sum_{\alpha=1}^{NF} c_\alpha \phi_\alpha,$$

where NF is the number of faces, c_α are the degrees of freedom computed by G_{RT}, and ϕ_α are the global basis functions for linear Raviart-Thomas finite element space. To conveniently mark local basis functions related to two particular elements (see Figure 1) we introduce the index-sets

$$\begin{aligned} \mathbb{I}_1 &= \{nlm, mlk, kln, nmk\}, && \text{indices to faces of element } T_{knlm}, \\ \mathbb{I}_2 &= \{mln, nlo, olm, mno\}, && \text{indices to faces of element } T_{omln}. \end{aligned}$$

Our goal is to minimize the quantity

$$\mathcal{J}(y) := \int\limits_{T_{knlm} \cup T_{omln}} \left(C(f + \operatorname{div} y)^2 + (y - \mathbf{A}\nabla v) \cdot (\mathbf{A}^{-1} y - \nabla v) \right) dx$$

by optimizing the degree of freedom c_{nlm} ($= c_{mln}$) shared by the two elments. Here

$C = (1+\beta)C_\Omega(1+\frac{1}{\beta})^{-1}$. This one-parametric problem is easily solved:

$$\frac{\partial \mathcal{J}}{\partial c_{nlm}} = 2 \int\limits_{T_{knlm}} \Big(C\Big(f + \sum_{\alpha\in\mathbb{I}_1} c_\alpha \mathrm{div}\, \phi_\alpha \Big) \mathrm{div}\, \phi_{nlm} - \phi_{nlm}\cdot\nabla v + \sum_{\alpha\in\mathbb{I}_1} c_\alpha \phi_\alpha \cdot \mathbf{A}^{-1}\phi_{nlm} \Big) dx +$$

$$+ 2 \int\limits_{T_{omln}} \Big(C\Big(f + \sum_{\alpha\in\mathbb{I}_2} c_\alpha \mathrm{div}\, \phi_\alpha \Big) \mathrm{div}\, \phi_{mln} - \phi_{mln}\cdot\nabla v + \sum_{\alpha\in\mathbb{I}_2} c_\alpha \phi_\alpha \cdot \mathbf{A}^{-1}\phi_{mln} \Big) dx = 0.$$

From the above we can solve a new value for the degree of freedom c_{nlm}:

$$c_{nlm} = \frac{A}{B}, \tag{9}$$

where

$$A =$$

$$\int\limits_{T_{knlm}} \Big(C\Big(f + \sum_{\alpha\in\mathbb{I}_1\setminus\{nlm\}} c_\alpha \mathrm{div}\, \phi_\alpha \Big) \mathrm{div}\, \phi_{nlm} - \phi_{nlm}\cdot\nabla v + \sum_{\alpha\in\mathbb{I}_1\setminus\{nlm\}} c_\alpha \phi_\alpha \cdot \mathbf{A}^{-1}\phi_{nlm} \Big) dx +$$

$$+ \int\limits_{T_{omln}} \Big(C\Big(f + \sum_{\alpha\in\mathbb{I}_2\setminus\{mln\}} c_\alpha \mathrm{div}\, \phi_\alpha \Big) \mathrm{div}\, \phi_{mln} - \phi_{mln}\cdot\nabla v + \sum_{\alpha\in\mathbb{I}_2\setminus\{mln\}} c_\alpha \phi_\alpha \cdot \mathbf{A}^{-1}\phi_{mln} \Big) dx,$$

and

$$B = - \int\limits_{T_{knlm}} \Big(C(\mathrm{div}\, \phi_{nlm})^2 + \phi_{nlm}\cdot\mathbf{A}^{-1}\phi_{nlm} \Big) dx - \int\limits_{T_{omln}} \Big(C(\mathrm{div}\, \phi_{mln})^2 + \phi_{mln}\cdot\mathbf{A}^{-1}\phi_{mln} \Big) dx.$$

We denote by P the procedure, which calculates the values of (9) for all degrees of freedom in a given mesh.

It should be noted, that the operator P can be applied to y as many times as wanted, and each time the value of $\mathcal{J}(y)$ decreases. In other words, the process is *monotone*. This post-processing method is also practical since it easily adapts parallelization.

4 NUMERICAL EXAMPLES

In this section, we test the performance of the error majorant $M_\oplus$ and indicator I with various methods of selecting y, which were derived in the previous section. For the purpose of measuring the performance of the majorant, we define the *efficiency index*

$$I_{eff} = \frac{M_\oplus}{|\!|\!| u - v |\!|\!|}.$$

The performance of the error indicator is tested by comparing the error distribution provided by the indicator to the exact error distribution.

To solve the model problem (1)-(2) we use the linear H^1 finite element. For the arbitary function y we use both the post-processing operators G_N, G_{RT}, and P and the global minimization method. For global minimization, we use the linear Raviart-Thomas finite element. In all numerical examples, the arbitary function y is computed on the same mesh on which the original numerical approximation v was computed.

Example 1:

$$\begin{aligned}&\Omega = [0,1]^2, \qquad f = 2(x_1(1-x_1) + x_2(1-x_2)),\\ &\mathbf{A} = \{a_{11} = a_{22} = 1, a_{12} = a_{21} = 0\}.\end{aligned}$$

For this problem the exact solution is known. Table 1 shows how the integral $\mathcal{J}(y)$ and the efficiency indexes I_{eff} for the upper bound $M_{\oplus}(v,y)$ behave with different y and different mesh-sizes. Post-processing methods G_N and G_{RT} fail to produce a flux that would satisfy the equilibrium condition, $\operatorname{div} y + f = 0$. For this reason, they do not provide a very accurate upper bound, and the values of I_{eff} are relatively large. By further post-processing, the value of the efficiency index can be decreased close to the one obtained by globally solved y_{glo}. According to numerical experiments, five iteration rounds are enough independent of the mesh size.

Example 2:

$$\begin{aligned}&\Omega = [0,1]^2, \qquad f = 2(10x_1(1-x_1) + x_2(1-x_2)),\\ &\mathbf{A} = \{a_{11} = 1, a_{22} = 10, a_{12} = a_{21} = 0\}.\end{aligned}$$

Also for this problem the exact solution is known. Figure 2 shows how the indicator I performs with different y in the second test example. Those elements, on which the error is greater than the average error, are marked with black color. In the top row, the leftmost picture is the exact error distribution. Here again y_{glo} denotes the function obtained by global minimization. As expected, global minimization of the upper bound gives good results. By using the operators G_N and G_{RT} we obtain good representations of error distributions. Moreover, further equilibration of $G_{RT}(\mathbf{A}\nabla v)$ by using the operator P does clearly improve the performance of I.

Example 3:

$$\begin{aligned}&\Omega = [0,2]^2,\\ &f = \begin{cases} 1 & \text{for } x_1 \in (0.5, 1.5) \\ 0 & \text{otherwise} \end{cases},\\ &\mathbf{A} = \begin{cases} \{a_{11} = 1, a_{22} = 1, a_{12} = a_{21} = 0\} & \text{for } x_1 \in (0.5, 1.5) \\ \{a_{11} = 10, a_{22} = 1, a_{12} = a_{21} = 0\} & \text{otherwise} \end{cases}.\end{aligned}$$

For this problem we do not know the exact solution. A reference solution was calculated in a very fine mesh to obtain a reference error distribution. From Figure 3 we see that this example is much more difficult compared to the previous example.

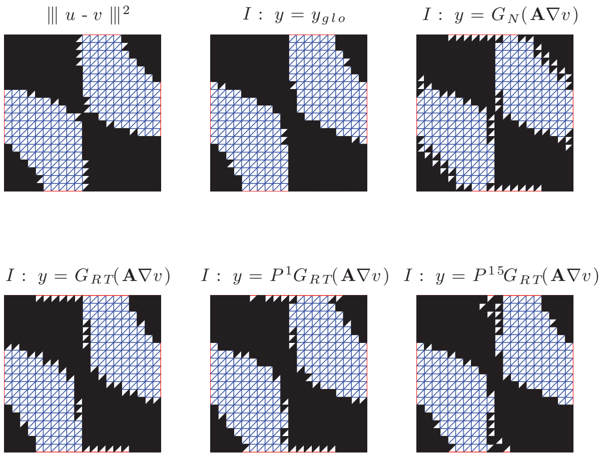


Figure 2: Performance of the post-processing operator P and the indicator I for Example 2.

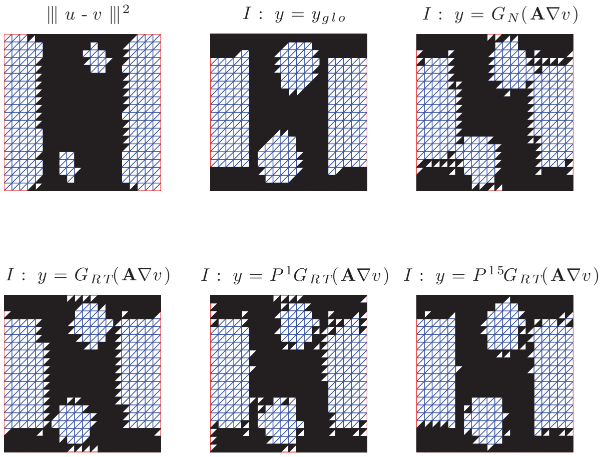


Figure 3: Performance of the post-processing operator P and the indicator I for Example 3.

	82 elems		1342 elems		8562 elems	
function y	$\mathcal{J}$	I_{eff}	$\mathcal{J}$	I_{eff}	$\mathcal{J}$	I_{eff}
$G_N(\mathbf{A}\nabla v)$	2.36e-3	2.46	2.48e-4	4.02	6.09e-5	6.53
$G_{RT}(\mathbf{A}\nabla v)$	3.24e-3	2.88	6.42e-4	8.35	2.00e-4	16.81
$P^1G_{RT}(\mathbf{A}\nabla v)$	2.40e-3	2.06	3.04e-4	3.80	8.10e-5	6.59
$P^2G_{RT}(\mathbf{A}\nabla v)$	2.17e-3	1.85	1.80e-4	2.25	3.93e-5	3.21
$P^5G_{RT}(\mathbf{A}\nabla v)$	2.07e-3	1.77	1.44e-4	1.79	2.34e-5	1.91
y_{glo}	2.00e-3	1.75	1.34e-4	1.72	2.06e-5	1.72

Table 1: Integral $\mathcal{J}(y)$ and efficiency index I_{eff} values with different mesh sizes and y for Example 1.

5 CONCLUSIONS

We conclude that in order to compute an efficient upper bound for the approximation error, the main problem is to obtain well enough equilibrated flux (minimizer of $\mathcal{J}(y)$). This task can be done with feasible computational effort using the presented new post-processing technique, which admits parallel processing.

For the purpose of obtaining the error distribution, the examples computed here do demonstrate some difference between the various post-processing methods tested. The averaging operators alone were able to represent the approximate flux well, but the proposed post-processing operator P was clearly shown to improve the quality of the approximate flux. As a natural consequence also the quality of error distributions was better after applying the post-processing operator P.

Acknowledgement. This research was supported by the grant N 40234 and N 40277 of the MASI Tekes Technology Programme and grant N 116895 of the Academy of Finland.

REFERENCES

[1] M. Ainsworth and J. T. Oden, A posteriori error estimation in finite element analysis, *Wiley ans Sons*, New York, (2000).

[2] I. Anjam, O. Mali, P. Neittaanmäki and S. Repin, A New Error Indicator for the Poisson Problem, In proceedings of the *10th Finnish Mechanics Days*, Jyväskylä, Finland, (2009).

[3] I. Babuška and R. Rodriguez, The problem of the selection of an a posteriori error indicator based on smoothing techniques, *Internat. J. Numer. Meth. Engrg.*, **36**, 539–567, (1993).

[4] F. Brezzi and M. Fortin, Mixed and hybrid finite element methods, *Springer Series in Computational Mathematics*, **15**, *New York*, (1991).

[5] P. Neittaanmäki and S. Repin, Reliable methods for computer simulation. Error control and a posteriori estimates, *Elsevier*, Amsterdam, (2004).

[6] P. A. Raviart and J. M. Thomas, Primal hybrid finite element methods for 2nd order elliptic equations, *Math. Comput.*, **31**, 391–413, (1977).

[7] S. Repin, A Posteriori Estimates for Partial Differential Equations, *Walter de Gruyter*, Berlin, (2008).

[8] O. C. Zienkiewicz and J. Z. Zhu, A simple error estimator and adaptive procedure for practical engineering analysis, *Internat. J. Numer. Meth. Engrg.*, **24**, 337–357, (1987).

PV

ON THE RELIABILITY OF ERROR INDICATION METHODS FOR PROBLEMS WITH UNCERTAIN DATA

by

I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin (2013)

A. Cangiani, R.L. Davidchack, E. Georgoulis, A.N. Gorban, J. Levesley, and M.V. Tretyakov (Eds.), Numerical Mathematics and Advanced Applications 2011, 811–819

Available online from Springer Link
http://link.springer.com/chapter/10.1007%2F978-3-642-33134-3_84.

On the reliability of error indication methods for problems with uncertain data

Immanuel Anjam, Olli Mali, Pekka Neittaanmäki, and Sergey Repin

Abstract This paper is concerned with studying the effects of uncertain data in the context of error indicators, which are often used in mesh adaptive numerical methods. We consider the diffusion equation and assume that the coefficients of the diffusion matrix are known not exactly, but within some margins (intervals). Our goal is to study the relationship between the magnitude of uncertainty and reliability of different error indication methods. Our results show that even small values of uncertainty may seriously affect the performance of all error indicators.

Acknowledgements This work was supported by the Finnish foundations Emil Aaltosen säätiö and KAUTE-säätiö, and the COMAS graduate school.

1 Introduction

In problems related to partial differential equations, it is usually assumed that data of the problem are known exactly. However, quite often the data at hand is not complete. In many problems, the data is uncertain within some intervals. Material functions, geometrical data, and boundary conditions may include uncertainties, which arise due to incomplete knowledge on the model.

Studying the effects caused by uncertain data gained the attention of researchers later than analysis of fully determined problems. The probabilistic approach is based in studying stochastic partial differential equations (see, e.g., [12]). Another approach (the so-called ”worst case scenario method”) has been analyzed in [5].

Immanuel Anjam, Olli Mali, Pekka Neittaanmäki
University of Jyväskylä, Department of Mathematical Information Technology, P.O. Box 35 (Agora), FI-40014 University of Jyväskylä, Finland, e-mail: immanuel.anjam@jyu.fi, olli.mali@jyu.fi, pekka.neittaanmaki@jyu.fi

Sergey Repin
V. A. Steklov Institute of Mathematics in St. Petersburg Fontanka 27, RU-191024, St. Petersburg, Russia, e-mail: repin@pdmi.ras.ru

In [6–8], two-sided estimates of the radius of the solution set were obtained for the reaction-diffusion problem. These estimates provide information on the *accuracy limit* generated by the uncertainty. These estimates are derived with the help of functional a posteriori estimates (for a consequent exposition of the theory see [9, 11]).

In this paper, we study the diffusion equation with uncertainty in the diffusion matrix. We assume that the uncertainty is of the form *mean value* ± *variation*, which is typical for engineering measurements.

Our goal is to study how incomplete knowledge about the coefficients of diffusion (physical parameters) impact the reliability of error indication. We have tested several commonly used indicators with the paradigm of a simple elliptic problem. The results show that the reliability of error indicators seriously depend on the rank of uncertainty encompassed in the diffusion matrix.

2 Problem definition and notation

Let $\Omega \in \mathbb{R}^d$ be a bounded and connected domain with Lipschitz continuous boundary $\partial\Omega$. By $H^1(\Omega)$ we denote the Sobolev space of scalar valued functions with square summable generalized derivatives. $H_0^1(\Omega)$ is the subspace of $H^1(\Omega)$ containing the functions vanishing on $\partial\Omega$. For vector valued functions, we also define the space $H(\mathrm{div},\Omega) := \{w \in L_2(\Omega,\mathbb{R}^d) \mid \mathrm{div}\, w \in L_2(\Omega)\}$.

We consider the simplest elliptic problem: find $u \in H_0^1(\Omega)$ such that

$$-\mathrm{div} A\nabla u = f \qquad \text{in } \Omega, \tag{1}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{2}$$

where $f \in L_2(\Omega)$. Assume that the coefficients are not fully known, i.e., the information that we really possess is that $A \in \mathscr{D}$, where

$$\mathscr{D} := \{A \in L_\infty(\Omega,\mathbb{M}^{d\times d}) \mid A = A_0 + \delta\Psi,\ \|\Psi\|_{L_\infty(\Omega,\mathbb{M}^{d\times d})} \leq 1\},$$

where $\mathbb{M}^{d\times d}$ is the space of symmetric matrices, $A_0 \in L_\infty(\Omega,\mathbb{M}^{d\times d})$ is the known "mean" matrix, and $\delta \geq 0$ is the magnitude of variations. In other words, we assume that A belongs to a set generated by limited perturbations of the "mean" data.

The generalized statement of (1)–(2) consists of finding $u \in H_0^1(\Omega)$ such that

$$\int_\Omega A\nabla u \cdot \nabla w\, dx = \int_\Omega f w\, dx, \qquad \forall w \in H_0^1(\Omega).$$

We assume that

$$\underline{c}|\xi|^2 \leq A_0\xi\cdot\xi \leq \overline{c}|\xi|^2,$$

where $0 < \underline{c} \leq \overline{c}$. Thus, the "mean" problem is elliptic and has a unique solution u_0. The condition

$$0 \leq \delta < \underline{c} \tag{3}$$

guarantees that the perturbed problem remains elliptic, and possesses a unique solution u with any $A \in \mathscr{D}$. With this condition the "solution mapping" $\mathscr{S} : \mathscr{D} \to H^1_0$ is well defined. The solution set generated by the uncertain data will be referred to as the set $\mathscr{S}(\mathscr{D}) \subset H^1_0$.

The standard L_2-norm $\|v\|_{2;\Omega}$ is denoted by $\|v\|_{\Omega}$. We also introduce the weighed L_2-norm for vector valued functions w:

$$\|w\|_{2,\mu;\Omega} = \|w\|_{\mu;\Omega} := \int_{\Omega} \mu w \cdot w dx.$$

Using this notation, for each $A \in \mathscr{D}$ we have the energy norm $\|\nabla v\|_{A;\Omega}$.

By $\mathscr{T}_h$ we denote the partition of the domain Ω to the union of non-overlapping triangles. An element in $\mathscr{T}_h$ is denoted by T. For any $T \in \mathscr{T}_h$, $\mathscr{E}(T)$ denotes edges of T, and $\mathscr{N}(T)$ the nodes. The sets

$$\mathscr{E}_h := \bigcup_{T \in \mathscr{T}_h} \mathscr{E}(T) \quad \text{and} \quad \mathscr{N}_h := \bigcup_{T \in \mathscr{T}_h} \mathscr{N}(T)$$

contain all edges and all nodes of $\mathscr{T}_h$, respectively. For the sake of convenience, we also define the set of edges which approximate the boundary of the domain by $\mathscr{E}_{h,\partial\Omega} := \{E \in \mathscr{E}_h \mid E \mathrel{\tilde{\subset}} \partial\Omega\}$. The sets

$$\omega_E := \bigcup_{E \in \mathscr{E}(T')} T' \quad \text{and} \quad \omega_X := \bigcup_{X \in \mathscr{N}(T')} T'$$

define patches of elements associated with a given edge $E \in \mathscr{E}_h$ and node $X \in \mathscr{N}_h$, respectively.

For every $E \in \mathscr{E}_h$, we assign a unit vector n_E, which it is orthogonal to E.

Henceforth, the symbol $|\cdot|$ is used to denote area of a domain or length of an edge. The number of elements in a set is denoted by $\#(\cdot)$ and the diameter of $T \in \mathscr{T}_h$ is denoted by h_T.

3 Error indication

In our analysis, we consider small disturbances of the matrix A of the form

$$A = A_0 + \delta B,$$

where the magnitude of variations δ satisfies the condition (3), and B is a symmetric 2×2 -matrix. We note that since the amount of matrixes contained in $\mathscr{D}$ is much larger than those representable in such a form, the sensitivity of error indicators with respect to data uncertainty is even higher than indicated on Table 1 and Fig. 1.

For each element $T \in \mathscr{T}_h$, the elements of B are chosen as follows:

$$B\big|_T = \begin{pmatrix} b_1 & b_3 \\ b_3 & b_2 \end{pmatrix}, \qquad b_1, b_2, b_3 \in \{-1, 0, 1\}, \qquad \forall T \in \mathscr{T}_h.$$

In other words, we generate a constant perturbation of magnitude δ in each element T. A perturbation generated in this way is clearly an extreme one. It suits our purposes, since we are trying to find a worst case situation that can occur with different diffusion matrices A which belong to the set $\mathscr{D}$.

Let $\mathbb{E}$ denote an error indicator computed on the mesh $\mathscr{T}_h$, which generates a set of non-negative numbers associated with elements, i.e.,

$$\mathbb{E}(A,u_h) := \{\mathbb{E}_T\}, \qquad \mathbb{E}_T \geq 0, \qquad \forall T \in \mathscr{T}_h.$$

Its input typically consists of the material data A, and a numerical solution u_h. The output is the vector $\{\mathbb{E}_T\}$, which contains an approximated error value $\mathbb{E}_T$ for each element T.

In computational practice, error indicators are used together with a *marker* $\Bbbk$ that marks elements (or other subdomains) where errors are excessively high. The function $\Bbbk$ takes as its input the vector produced by an error indicator $\mathbb{E}$, and returns a boolean function indicating by 1 the elements which are to be refined, and by 0 all other elements. The output of $\Bbbk(\mathbb{E})$ is essentially the list of those elements T, which contain the majority of the error (according to the indicator used). We refer to the boolean output of $\Bbbk$ as a *marking*. The marker can, for example, choose to mark some percentage of the elements (”bulk criterion”), or to mark those elements whose indicator value is greater than the average of all the values. In this short note we confine ourselves to the case where $\Bbbk$ marks a certain amount (N_{ref}) of elements, where the highest values of errors have been indicated.

Our analysis of effects caused by data uncertainty is based on the following method. Let $\mathbb{E}$ be the indicator to test. We select a mesh $\mathscr{T}_h$ and select a certain amount of matrices $A_j = A_0 + \delta B_j$ for some given δ (uncertainty parameter). For each exact solution $u_j = \mathscr{S}(A_j)$, we compute the corresponding approximations u_{jh} on the mesh $\mathscr{T}_h$. Then, for each u_{jh}, we calculate the error indicator $\mathbb{E}_j = \mathbb{E}(A_j, u_{jh})$, and the corresponding markings $\Bbbk(\mathbb{E}_j)$.

The difference of two markings is given by the boolean measure

$$\mathrm{diff}(\Bbbk, \mathbb{E}_i, \mathbb{E}_j) := 1 - \frac{\sum(\Bbbk(\mathbb{E}_i) \wedge \Bbbk(\mathbb{E}_j))}{N_{ref}} \in [0,1],$$

where $\wedge$ is the logical multiplication operator. If $\mathrm{diff}(\Bbbk, \mathbb{E}_i, \mathbb{E}_j) = 0$, then small variations of the data do not affect the process of marking. In opposite, if $\mathrm{diff}(\Bbbk, \mathbb{E}_i, \mathbb{E}_j)$ is close to one, then the lists of elements selected for refinement by $\mathbb{E}_i$ and $\mathbb{E}_j$ are quite different.

The maximal difference between all markings is given by the quantity

$$\Theta := \max_{i,j}\{\mathrm{diff}(\Bbbk, \mathbb{E}_i, \mathbb{E}_j)\},$$

which shows the maximal difference produced by an error indicator with different diffusion matrices from the set $\mathscr{D}$.

From now on, we will denote by u_h an approximation of (1)–(2) calculated with the help of standard linear Courant elements.

We have tested the following six most commonly used error indicators.

Indicators based on averaging. The well known node averaging indicator (see, e.g., [13, 14]) reads

$$\mathbb{E}_{X,T} := \|G_X u_h - A\nabla u_h\|_{A^{-1};T}, \tag{4}$$

and a similar indicator (we call the edge averaging indicator, see, e.g., [11]) reads

$$\mathbb{E}_{E,T} := \|P^p(G_E u_h) - A\nabla u_h\|_{A^{-1};T}. \tag{5}$$

The averaging operators G_X and G_E are defined by the relations

$$G_X u_h(X) = \sum_{T\in\omega_X} \frac{|T|}{|\omega_X|} (A\nabla u_h)\big|_T \quad \text{and} \quad G_E u_h(E) = \frac{|E|}{\#\omega_E} \sum_{T\in\omega_E} (A\nabla u_h)\big|_T \cdot n_E,$$

which define the values of $G_X u_h$ and $G_E u_h$ at the node X and edge E, respectively. Then, the averaged function $G_X u_h$ is defined by piecewise affine extension, and $G_E u_h$ by extension with the help of linear Raviart-Thomas elements (see, e.g., [10]). In (5), the operator P is a post-processing operator, which produces more accurate approximations for the exact flux $A\nabla u$ by minimizing the residual $\|f + \operatorname{div} y\|^2_{\omega_E}$ on all subdomains ω_E (see, e.g., [1,11]). Here $y \in H(\operatorname{div},\Omega)$ is a vector valued function generated by the averaging operator G_E (we assume that $P^0(G_E u_h) = G_E u_h$).

Residual based indicators. Residual based error indicators form the class of mostly used error indicators (see, e.g., [2, 13]). We consider the standard residual error indicator

$$\mathbb{E}_{RF,T} = \left(h_T^2 \|f_T\|_T^2 + \frac{1}{2} \sum_{E\in\mathscr{E}_h(T)\setminus\mathscr{E}_{h,\partial\Omega}} |E|\, \|[n_E \cdot A\nabla u_h]_E\|_E^2 \right)^{1/2}, \tag{6}$$

and the indicator containing only jump terms

$$\mathbb{E}_{RJ,T} = \left(\frac{1}{2} \sum_{E\in\mathscr{E}_h(T)\setminus\mathscr{E}_{h,\partial\Omega}} |E|\, \|[n_E \cdot A\nabla u_h]_E\|_E^2 \right)^{1/2}. \tag{7}$$

Here f_T denotes the mean value of f on T, i.e., $f_T := \frac{1}{|T|}\int_T f\,dx$.

Global averaging indicator. The global averaging indicator (see, e.g., [3,4]) reads

$$\mathbb{E}_{GA,T} := \|y_{GA} - A\nabla u_h\|_{A^{-1};T}, \tag{8}$$

where y_{GA} is calculated by global minimization of $\|y_{GA} - A\nabla u_h\|^2_{A^{-1};\Omega}$. This minimization procedure results in the problem: find $y_{GA} \in H(\operatorname{div},\Omega)$ such that

$$\int_\Omega A^{-1} y_{GA} \cdot w\,dx = \int_\Omega \nabla u_h \cdot w\,dx, \qquad \forall w \in H(\operatorname{div},\Omega).$$

In our tests we used linear Raviart-Thomas finite elements (see, e.g., [10]) in order to find globally averaged indicator on the mesh $\mathscr{T}_h$.

Error indicator generated by the functional type error majorant. The difference between the exact solution u and an approximation u_h is bounded from above by the functional error majorant $M_\oplus$ (see, e.g., [9, 11]):

$$\|\nabla(u-u_h)\|^2_{A;\Omega} \leq C_1 \|f + \mathrm{div} y_F\|^2_\Omega + C_2 \|y_F - A\nabla u_h\|^2_{A^{-1},\Omega} := M_\oplus(A, u_h, y_F),$$

where $C_1 = (1+\alpha)C^2_\Omega \underline{c}^{-1}$ and $C_2 = (1+\alpha^{-1})$. The constant C_Ω is the Friedrich's constant. The above inequality holds for all $y_F \in H(\mathrm{div}, \Omega)$ and $\alpha \in \mathbb{R}_+$. The latter term in the upper bound $M_\oplus$ can be used as an error indicator (see [11] for the mathematical justification of this indicator):

$$\mathbb{E}_{F,T} := \|y_F - A\nabla u_h\|_{A^{-1};T}. \tag{9}$$

The function y_F is calculated by minimization of $M_\oplus$. This minimization procedure results in a problem for $y_F \in H(\mathrm{div}, \Omega)$ and $\alpha \in \mathbb{R}_+$:

$$\int_\Omega \big(C_1 \mathrm{div} y_F\, \mathrm{div} w + C_2 A^{-1} y_F \cdot w\big) dx = \int_\Omega (C_2 \nabla u_h \cdot w - C_1 f\, \mathrm{div} w) dx, \ \forall w \in H(\mathrm{div}, \Omega).$$

This problem was also solved with the help of linear Raviart-Thomas finite elements.

4 Numerical results and conclusions

Approximate solutions of the problem (1)–(2) have been computed using standard Courant type finite element approximations. Indicators (8) and (9) were calculated with the help of linear Raviart-Thomas finite elements. All the problems were calculated on same regular meshes, and systems of linear simultaneous equations were solved by exact methods. In view of this fact, approximate solutions possess Galerkin orthogonality property, and, therefore, the residual error indicator (6) can be used. For the edge averaging indicator (5), we set $p = 5$ (the amount of times P is applied). All calculations were performed with the `MATLAB` computing environment on a 64 processor SMP server with 1 TB of RAM.

In total, a mesh contains $N_{elem} := \#\mathcal{T}_h$ elements. Since in this paper we calculate approximations of (1)–(2) using linear Courant elements, the amount of degrees of freedom N_{dof} of an approximation equals the number of nodes $\#\mathcal{N}_h$. We chose to mark 30% of elements of a mesh to be refined, i.e., $N_{ref} = 0.3 \times N_{elem}$.

We studied how the magnitude of variations δ affects error indicators, and discuss typical results with the paradigm of a simple problem where

$$\Omega = [0,1]^2, \qquad A_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \text{and} \quad f = 2(x_1(1-x_1) + x_2(1-x_2)). \tag{10}$$

The exact solution of this problem is $u_0 = x_1(1-x_1)x_2(1-x_2)$.

Using the procedure explained in Sect. 3, we have tested six different indicators for six different meshes. The results are exposed on Table 1 and Fig. 1. It is worth to outline again, that the actual sensitivity of error indicators with respect to data uncertainty is even higher than in the results reported below (which should be viewed as lower bounds of the true sensitivity).

Table 1 shows how the values of Θ (associated with the indicators (4)–(9)) depend on the amount of elements N_{elem} (or amount of degrees of freedom N_{dof}) and the parameter δ. It is easy to see that sufficiently small values of Θ (which correspond to relatively stable performance of an error indicator) are obtained only for small δ (such as 0.005 or 0.01) and rather moderate amount of elements. If values of δ are not very small (e.g., 0.04) then all indicators may generate quite different markings. We recall that $\Theta = 1$ if indicators computed for different elements of the solution set $\mathscr{D}$ may generate completely opposite markings.

A selection of these numbers are presented on Fig. 1 in a graphical way, which allows us to compare different indicators with each other. We conclude that even in this very simple problem small uncertainties in the matrix coefficients may seriously corrupt the process of error indication. This phenomenon does not depend on a particular error indicator. Actually, it shows that in real life computations error indication procedures (and subsequent mesh refinement) cannot be performed without an adequate analysis of data uncertainty.

Table 1 The values of Θ with two digit accuracy (example (10)).

(a) $\boxed{\mathbb{E}_X}$ node averaging (4)

N_{elem}	N_{dof}	δ					
		0.005	0.01	0.02	0.03	0.04	0.05
800	441	0.15	0.28	0.52	0.68	0.80	0.85
3200	1681	0.30	0.53	0.80	0.89	0.88	0.98
12800	6561	0.53	0.80	0.88	1	1	1
51200	25921	0.80	0.88	1	1	1	1
115200	58081	0.89	1	1	1	1	1

(b) $\boxed{\mathbb{E}_E}$ edge averaging (5)

N_{elem}	N_{dof}	δ					
		0.005	0.01	0.02	0.03	0.04	0.05
800	441	0.16	0.27	0.50	0.67	0.75	0.84
3200	1681	0.28	0.51	0.77	0.87	0.89	0.87
12800	6561	0.51	0.77	0.89	0.95	1	1
51200	25921	0.77	0.89	1	1	1	1
115200	58081	0.88	0.96	1	1	1	1

(c) $\boxed{\mathbb{E}_{RF}}$ residual, full (6)

N_{elem}	N_{dof}	δ					
		0.005	0.01	0.02	0.03	0.04	0.05
800	441	0.28	0.41	0.65	0.80	0.90	0.94
3200	1681	0.42	0.64	0.89	0.95	0.96	0.96
12800	6561	0.65	0.89	0.96	0.97	1	1
51200	25921	0.89	0.96	1	1	1	1
115200	58081	0.95	0.97	1	1	1	1

(d) $\boxed{\mathbb{E}_{RJ}}$ residual, jump (7)

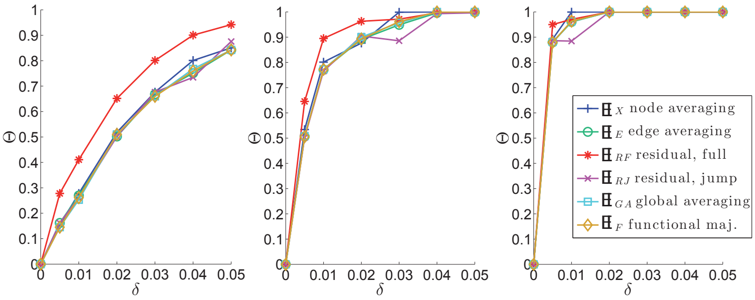
N_{elem}	N_{dof}	δ					
		0.005	0.01	0.02	0.03	0.04	0.05
800	441	0.16	0.26	0.50	0.68	0.73	0.88
3200	1681	0.28	0.51	0.77	0.90	0.91	0.89
12800	6561	0.50	0.77	0.90	0.89	0.99	1
51200	25921	0.76	0.90	0.99	1	1	1
115200	58081	0.89	0.88	1	1	1	1

(e) $\boxed{\mathbb{E}_{GA}}$ global averaging (8)

N_{elem}	N_{dof}	δ					
		0.005	0.01	0.02	0.03	0.04	0.05
800	441	0.15	0.25	0.51	0.66	0.77	0.84
3200	1681	0.29	0.51	0.78	0.89	0.91	0.88
12800	6561	0.51	0.77	0.90	0.96	1	1
51200	25921	0.77	0.89	1	1	1	1
115200	58081	0.88	0.96	1	1	1	1

(f) $\boxed{\mathbb{E}_F}$ functional maj. (9)

N_{elem}	N_{dof}	δ					
		0.005	0.01	0.02	0.03	0.04	0.05
800	441	0.15	0.26	0.51	0.66	0.76	0.84
3200	1681	0.28	0.51	0.78	0.89	0.89	0.87
12800	6561	0.51	0.77	0.89	0.96	1	1
51200	25921	0.77	0.89	1	1	1	1
115200	58081	0.88	0.96	1	1	1	1



(a) $N_{elem} = 800$
$N_{dof} = 441$

(b) $N_{elem} = 12800$
$N_{dof} = 6561$

(c) $N_{elem} = 115200$
$N_{dof} = 58081$

Fig. 1 The values of Θ plotted against the magnitudes of variation δ, for three meshes (example (10)).

References

1. Anjam, I., Mali, O., Neittaanmäki, P., Repin, S.: A new error indicator for the Poisson problem. In: Mäkinen, R., Neittaanmäki, P., Tuovinen, T., Valpe, K. (eds.), Proceedings of the 10th Finnish Mechanics Days, pp. 324–330 (2009)
2. Babuška, I., Rheinboldt W.C.: Error estimates for adaptive finite element computations. SIAM J. Numer. Anal. **15**, pp. 736–754 (1978)
3. Bartels, S., Carstensen, C.,: Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. Part II: Higher order FEM. Math. Comput. **71**(239), pp. 971–994 (2002)
4. Carstensen, C., Bartels, S.: Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. Part I: Low order conforming, nonconforming, and mixed FEM. Math. Comput. **71**(239), pp. 945–969 (2002)
5. Hlaváček, I., Chleboun, J., Babuška, I.: Uncertain input data problems and the worst scenario method. Elsevier, Amsterdam (2004)
6. Mali, O., Repin, S.: Estimates of accuracy limit for elliptic boundary value problems with uncertain data. Adv. Math. Sci. Appl. **19**(2), pp. 525–537 (2009)
7. Mali, O., Repin, S.: Estimates of the indeterminacy set for elliptic boundary value problems with uncertain data. J. Math. Sci. **150**(1), pp. 1869–1874 (2008)
8. Mali, O., Repin, S.: Two-sided estimates of the solution set for the reaction-diffusion problem with uncertain data. Comput. Methods Appl. Sci. **15**, pp. 183–198 (2010)
9. Neittaanmäki, P., Repin, S.: Reliable methods for computer simulation. Error control and a posteriori estimates. Elsevier, Amsterdam (2004)
10. Raviart, P.A., Thomas, J.M.: Primal hybrid finite element methods for 2nd order elliptic equations. Math. Comput. **31**, pp. 391–413 (1977)
11. Repin, S.: A posteriori estimates for partial differential equations. Walter de Gruyter, Berlin (2008)
12. Schuëller, G.I.: A state-of-the-art report on computational stochastic mechanics. Prob. Engrg. Mech. **12**(4), pp. 197–321 (1997)
13. Verfürth, R.: A review of a posteriori error estimation and adaptive mesh-refinement techniques. Wiley and Sons, Teubner-Verlag, New York (1996)
14. Zienkiewicz, O.C., Zhu, J.Z.: A simple error estimator and adaptive procedure for practical engineering analysis. Int. J. Numer. Meth. Engrg. **24**, pp. 337–357 (1987)

PVI

A UNIFIED APPROACH TO MEASURING ACCURACY OF ERROR INDICATORS

by

I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin (2014)

W. Fitzgibbon, Yu. A. Kuznetsov, P. Neittaanmaki, and O. Pironneau (Eds.),
Modeling, Simulation and Optimization for Science and Technology (to appear)

Chapter 1
A Unified Approach to Measuring Accuracy of Error Indicators

Immanuel Anjam, Olli Mali, Pekka Neittaanmäki, and Sergey Repin

Abstract In this paper, we present a unified approach to error indication for elliptic boundary value problems. We introduce two different definitions of the accuracy (weak and strong) and show that various indicators result from one principal relation. In particular, this relation generates all the main types of error indicators, which have already gained high popularity in numerical practice. Also, we discuss some new forms of indicators that follow from a posteriori error majorants of the functional type and compare them with other indicators. Finally, we discuss another question related to accuracy of error indicators for problems with incompletely known data.

1.1 Introduction

Error indicators play an important role in mesh-adaptive numerical algorithms, which currently dominate in mathematical and numerical modelling of various models in physics, chemistry, biology, economics, and other sciences. Their goal is to present a comparative measure of errors related to different parts of the computational domain, which could suggest a reasonable way of improving the finite–dimensional space used to compute the approximate solution. An "ideal" error indicator must possess several properties: efficiency, computability, and universality. In

I. Anjam, O. Mali, P. Neittaanmäki, and S. Repin
Department of Mathematical Information Technology, P.O. Box 35 (Agora), 40014 University of Jyväskylä, Finland, e-mail: `immanuel.anjam@jyu.fi`

O. Mali
e-mail: `olli.mali@jyu.fi`

P. Neittaanmäki
e-mail: `pekka.neittaanmaki@jyu.fi`

S. Repin
e-mail: `sergey.repin@jyu.fi`

other words, it must correctly reproduce the distribution of errors, be indeed computable, and be applicable to a wide set of approximations. In practice, it is very difficult to satisfy all these requirements simultaneously so that different error indicators are focused on different aims and stress some properties at the sacrifice of others. We discuss the mathematical origins and algorithmic implementation of the most frequently used error indicators. The literature devoted to this subject is vast.

Indicators based upon post–processing (e.g., averaging) of numerical solutions are among the most widely used. Among first publications in this direction we mention [54, 52], which generated an interest in gradient recovery methods. Similar methods were investigated in numerous publications (see, e.g., [2, 8, 7, 11, 27, 47, 49, 50, 51, 53]). Mathematical justifications of the error indicators obtained in this way follow from the *superconvergence* phenomenon (see, e.g., [32, 48, 30]). Post–processing based on global averaging procedures can be performed under weaker assumptions, which makes them applicable to a wider class of problems (see, e.g., [18, 19, 27]). Another class of post–processing methods generate equilibrated (or almost equilibrated) fluxes (see, e.g., [3, 16, 33]).

Residual type error indicators is another wide class of indicators. They originate from the papers [5, 6]. Various modifications and advanced forms have been discussed in numerous publications (see, e.g., [2, 3, 8, 9, 17, 20, 26, 29, 47, 24]).

Runge type indicators are based on the solutions on an enriched set of basis functions. A special class are the hierarchical error indicators, which are constructed with the help of auxiliary problems on enriched finite-dimensional subspaces (local or global) (see, e.g., [1, 22, 25, 23] and the references therein).

Evaluation of approximation errors in terms of special “goal-oriented” quantities is very popular in engineering computations. A consequent exposition can be found in [10] and in numerous publications devoted to *goal-oriented* a posteriori error estimates and applications of them to various problems (see, e.g, [13, 14, 28, 35, 31, 36, 37, 39, 40, 45, 46]).

The outline of the paper is as follows. In Sect. 1.2 we define strong and weak accuracy measures for error indicators. Section 1.3 presents a unified conception of error indicators which contains all main types of error indicators used in practice. Section 1.4 contains numerical tests, which show the performance of various error indicators applied to finite element approximations of boundary value problem in a domain with reentrant corners and jumping coefficients. We come to the conclusion that error indication of some zones containing excessively high errors is properly done by all error indicators. However, quantitative results are quite different and some indicators seriously overestimate true values of the error. In Sect. 1.5 we discuss the effects that the incompletely known data has on the applicability of error indicators and present related numerical examples.

1.2 Error Indicators for FEM Solutions

Let T_s, $s = 1, 2, ..., N$, be elements (subdomains) associated with the mesh $\mathcal{T}_h$ (with characteristic size h), and let u_h be an approximate solution computed on this mesh. Henceforth, the corresponding finite dimensional space is denoted by V_h, so that $u_h \in V_h$. Then, the true error is $e = u - u_h$. Denote by $m_s(e)$ the value of the error measure m associated with T_s. Usually, the error measure $m_s(e)$ is defined as a certain integral of $u - u_h$ related to T_s. For example, local error measures of approximate solutions to linear elliptic problems are often presented by the integrals

$$\left(\int_{T_s} |u - u_h|^2 dx \right)^{\frac{1}{2}} \quad \text{or} \quad \left(\int_{T_s} |\nabla(u - u_h)|^2 dx \right)^{\frac{1}{2}}.$$

The components of the vector

$$\mathbf{m}(e) = \{m_1(e), m_2(e), ..., m_N(e)\}$$

are nonnegative numbers, which may be rather different.

If the overall error encompassed in u_h is too big, then a new approximate solution should be computed on a new (refined) mesh $\mathcal{T}_{h_{\rm ref}}$. Comparative analysis of $m_s(e)$ suggests an idea where to add new degrees of freedom (new trial functions). However, in real life computations the vector $\mathbf{m}(e)$ is not known and, therefore, an error indicator $\mathbb{E}(u_h)$ is used. The corresponding approximate values of errors $\mathbb{E}_s$ associated with the elements form the vector

$$\mathbb{E}(u_h) = \{\mathbb{E}_1, \mathbb{E}_2, ..., \mathbb{E}_N\},$$

which is used instead of $\mathbf{m}(e)$.

If the vector $\mathbb{E}(u_h)$ is close to $\mathbf{m}(e)$, i.e.,

$$\mathbf{m}(e) \approx \mathbb{E}(u_h), \tag{1.1}$$

then a new mesh $\mathcal{T}_{h_{\rm ref}}$ can be efficiently constructed on the basis of comparative analysis of $\mathbb{E}_s$. However, the fact that the adaptive procedure is efficient depends on how accurately the condition (1.1) is satisfied and how efficiently the information encompassed in $\mathbb{E}(u_h)$ is used to improve approximations.

Certainly, the condition (1.1) looks vague unless a formal definition of the sign $\approx$ is given. Despite the huge amount of publications focused on error indication, to the best of our knowledge no commonly used definition has yet been accepted. Different authors may claim (explicitly or implicitly) different things, so the words "good error indicator" may take on a variety of meanings.

Below we suggest definitions, which can be used for a reasonable qualification of error indicators. They define "strong" and "weak" meanings of $\approx$, respectively.

Definition 1.1. The indicator $\mathbb{E}(u_h)$ is ε-accurate on the mesh $\mathcal{T}_h$ if

$$\mathscr{M}(\mathbb{E}(u_h)) := \frac{|\mathbf{m}(e) - \mathbb{E}(u_h)|}{|\mathbf{m}(e)|} \leq \varepsilon. \tag{1.2}$$

The value of $\mathscr{M}(\mathbb{E}(u_h))$ is the strongest quantitative measure of the accuracy of $\mathbb{E}(u_h)$.

This definition imposes strong requirements on $\mathbb{E}(u_h)$. Indeed, (1.2) guarantees that inaccuracies in the error distribution computed by $\mathbb{E}(u_h)$ are much smaller (provided that ε is a small number) than the overall error. Therefore, an indicator should be regarded as "accurate", if it meets (1.2) with relatively coarse ε.

From (1.2) it follows that the so-called efficiency index

$$I_{\text{eff}}(\mathbb{E}(u_h)) := \frac{|\mathbb{E}(u_h)|}{|\mathbf{m}(e)|} \leq 1 + \mathscr{M}(\mathbb{E}(u_h)) \tag{1.3}$$

is close to 1, which means that $|\mathbb{E}(u_h)|$ provides a good evaluation of the overall error $|\mathbf{m}(e)|$.

The efficiency of $\mathbb{E}(u_h)$ may be different for different meshes and approximate solutions. It is desirable that the indicator is accurate for a sufficiently wide class of approximations and meshes. The wider class of approximations served by an indicator, the better it is from the computational point of view.

The majority of indicators suggested for finite element approximations are applicable only to Galerkin approximations (or to approximations that are very close to Galerkin solutions). Properties of the mesh used are also very important, and theoretical estimates of the quality of error indicators usually involve constants dependent on the aspect ratio of finite elements.

In adaptive finite element schemes, subsequent approximations are often constructed on nested meshes, where a refined mesh is obtained by "splitting" elements (h-refinement) or by increasing the amount and order of basis functions (p-refinement) of the current mesh.

A detailed discussion on refinement methods can be found in, e.g., [4, 21].

Typical adaptive schemes consists of solving the problem several times on a sequence of improving subspaces. In this type of practice, error indicators are used together with a *marker* that marks elements (subdomains) where errors are excessively high. A new subspace $V_{h_{\text{ref}}}$ is constructed in such a way that these errors are diminished.

Let $\mathbf{B}$ denote the Boolean set $\{0,1\}$ (we can assign the meaning "NO" to 0 and "YES" to 1). By $\mathbf{B}^N$ we denote the set of Boolean valued arrays (associated with one-, two- or multidimensional meshes) of total length N. If $\mathbf{b} = \{\flat_1, \flat_2, ..., \flat_N\} \in \mathbf{B}^N$, then $\flat_s \in \mathbf{B}$ for any $s = 1, 2, ...N$. It is assumed that in the new mesh the elements (subdomains) marked by 1 should be refined, while those marked by 0 should be preserved (see Fig. 1.1). Note that the refined mesh in Fig. 1.1 contains the so-called "hanging nodes". In order to avoid them it is often necessary to refine also some neighboring subdomains marked by 0.

Remark 1.1. Modern mesh adaptation algorithms often make coarsening of a mesh in subdomains where local errors are insignificant (see, e.g., [15, 43, 44, 38, 12,

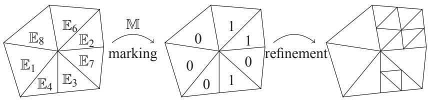


Fig. 1.1 Marking procedure and a refined mesh

31, 42] and the references cited therein). In this case, elements of $\mathbf{B}^N$ may attain three values: $\{-1,0,1\}$. The elements marked by -1 should be further aggregated in larger blocks.

From the mathematical point of view, marking is an operation performed by a special operator.

Definition 1.2. Marker $\mathbb{M}$ is a mapping (operator) acting from the set $\mathbb{R}_+^N$ (which contains estimated values of local errors) to the set $\mathbf{B}^N$.

Different markers generate different selection procedures, which are applied to the array of errors evaluated by an indicator $\mathbb{E}(u_h)$ in order to obtain a boolean array **b**. Further refinement is performed with the help of data encompassed in **b**. To compare different error indicators in the context of elementwise marking, we introduce two operations with Boolean valued arrays. Let $\mathbf{a}=\{a_i\}$ and $\mathbf{b}=\{b_i\}$ be elements of $\mathbf{B}^N$. By $[\![\mathbf{a}]\!]$ we denote the sum $\sum_{i=1}^N a_i$ and $\equiv$ denotes the componentwise logical equivalence rule, i.e.,

$$\{\mathbf{a}\equiv\mathbf{b}\}_i=\begin{cases}1 & \text{if } a_i=b_i\\ 0 & \text{if } a_i\neq b_i.\end{cases}$$

Definition 1.3. An indicator $\mathbb{E}(u_h)$ is ε-accurate on the mesh $\mathscr{T}_h$ with respect to the marker $\mathbb{M}$ if

$$\mathscr{M}(\mathbb{E}(u_h),\mathbb{M}):=1-\frac{[\![\mathbb{M}(\mathbf{m}(e))\equiv\mathbb{M}(\mathbb{E}(u_h))]\!]}{N}\leq\varepsilon. \tag{1.4}$$

It is easy to see that the accuracy measure $\mathscr{M}(\mathbb{E}(u_h),\mathbb{M})$ is much weaker than the measure introduced in Definition 1.1.

1.3 General Scheme for Deriving Error Indicators

Practically all known error indicators can be suggested within the framework of a unified scheme suggested in [34], where it is discussed with the paradigm of the Poisson equation. In this section, we present a generalized version of this scheme, which is applicable to a wide spectrum of elliptic type problems. Namely, we consider the class of boundary value problems

$$\Lambda^*\mathscr{A}\Lambda u+\mathscr{B}u=f \quad \text{in } \Omega, \quad f\in\mathscr{V}, \tag{1.5}$$

$$u=u_0 \quad \text{on } \Gamma, \tag{1.6}$$

where Ω is an open bounded connected subset in $\mathbb{R}^d$ with Lipschitz continuous boundary Γ. Here, $\mathscr{V}$ and U are two Hilbert spaces with the inner products by $(\cdot,\cdot)_{\mathscr{V}}$ and $(\cdot,\cdot)_U$ respectively. These products generate the norms $\|\cdot\|_{\mathscr{V}}$ and $\|\cdot\|_U$. The operator $\mathscr{A} : U \to U$ and $\mathscr{B} : \mathscr{V} \to \mathscr{V}$ are linear, self-adjoint, and positive definite operators. $\Lambda : V \to U$ is a bounded linear operator, $V \subset \mathscr{V}$ is a Hilbert space generated by the inner product $(w,v)_V := (w,v)_{\mathscr{V}} + (\Lambda w, \Lambda v)_U$. Henceforth, V_0 denotes a convex, closed and non-empty subspace of V such that $V_0 \subset V \subset \mathscr{V} \subset V_0^*$.

Typically, V is a Sobolev space associated with the differential operator Λ and V_0 contains the functions, which satisfy homogeneous Dirichlet boundary conditions on a part of the boundary. We consider boundary value problems associated with energy functionals of the form:

$$J(w) := \tfrac{1}{2}(\mathscr{A}\Lambda w, \Lambda w)_U + \tfrac{1}{2}(\mathscr{B}w, w)_{\mathscr{V}} - (f,w)_{\mathscr{V}}, \tag{1.7}$$

where $f \in \mathscr{V}$. We assume that

$$(\mathscr{A}y, y)_U \geq c_1 \|y\|_U^2 \quad \forall y \in U, \tag{1.8}$$

and

$$\|w\|_{\mathscr{V}} \leq C_F \|\Lambda w\|_U, \quad \forall w \in V_0. \tag{1.9}$$

The adjoint operator $\Lambda^* : U \to V_0^*$ is defined by the relation

$$\langle \Lambda^* y, w \rangle = (y, \Lambda w)_U, \quad \forall y \in U,\ w \in V_0, \tag{1.10}$$

where $\langle \cdot,\cdot \rangle$ denotes the pairing of V_0 and its conjugate V_0^* and $\langle \Lambda^* y, w \rangle$ is the value of the functional $\Lambda^* y \in V_0^*$ at $w \in V_0$.

Let $a : V_0 \times V_0 \to \mathbb{R}$ denote the symmetric bilinear form

$$a(u,w) := (\mathscr{A}\Lambda u, \Lambda w)_U + (\mathscr{B}u, w)_{\mathscr{V}}. \tag{1.11}$$

Under the above made assumptions, the form a is V-elliptic and defines the energy norm $|||w||| := \sqrt{a(w,w)}$. We define additional equivalent norms in U

$$\|y\|_{\mathscr{A}}^2 := (\mathscr{A}y, y)_U \quad \text{and} \quad \|y\|_{\mathscr{A}^{-1}}^2 := (\mathscr{A}^{-1}y, y)_U.$$

Now

$$J(w) := \frac{1}{2}a(w,w) - (f,w)_{\mathscr{V}} \tag{1.12}$$

and the (generalized) solution u is the minimizer of the variational problem

$$J(u) = \min_{w \in V_0} J(w). \tag{1.13}$$

By standard arguments, it is easy to prove that the minimizer exists and is unique. Moreover, it satisfies the relation

$$a(u,w) = (f,w)_{\mathscr{V}}, \quad \forall w \in V_0, \tag{1.14}$$

which presents a generalized solution of (1.5)–(1.6).

Note that

$$\sup_{w\in V_0}\left\{(\mathscr{A}\Lambda(u-v),\Lambda w)_U+(\mathscr{B}(u-v),w)_{\mathscr{V}}-\frac{1}{2}a(w,w)\right\}\leq$$
$$\leq\sup_{\tau\in U}\left\{(\mathscr{A}\Lambda(u-v),\tau)_U-\frac{1}{2}(\mathscr{A}\tau,\tau)_U\right\}+\sup_{\eta\in\mathscr{V}}\left\{(\mathscr{B}(u-v),\eta)_{\mathscr{V}}-\frac{1}{2}(\mathscr{B}\eta,\eta)_{\mathscr{V}}\right\}=$$
$$=\frac{1}{2}\,|||\,u-v\,|||^2\,.$$

On the other hand,

$$\sup_{w\in V_0}\left\{a(u-v,w)-\frac{1}{2}\,|||\,w|||^2\right\}\geq\frac{1}{2}\,|||\,e\,|||^2\,.$$

Thus,

$$|||e|||^2=\sup_{w\in V_0}\left\{-\,|||\,w\,|||^2-2\ell_v(w)\right\},\tag{1.15}$$

where $\ell_v(w):=(\mathscr{A}\Lambda v,\Lambda w)_U+(\mathscr{B}v,w)_{\mathscr{V}}-(f,w)_{\mathscr{V}}$ is the *residual functional*. It is easy to show that the variational problem on the right-hand side of (1.15) has a unique solution and this solution is $w=u-v$. Indeed,

$$\ell_v(u-v)=(\mathscr{A}\Lambda v,\Lambda(u-v)-\mathscr{A}\Lambda u,\Lambda(u-v))_U+(\mathscr{B}v,u-v)_{\mathscr{V}}-(\mathscr{B}u,u-v)_{\mathscr{V}}$$
$$=-\,|||\,e|||^2,$$

and we see that the right-hand side coincides with the left-hand one. Hence, (1.15) implies the relation

$$|\ell_v(e)|=|||e\,|||^2\,.$$

We can use (1.15) to deduce computable error indicators in the following three principal ways:

1. Assume that we can estimate the residual functional from above as follows:

$$\ell_v(w)\leq\overline{M}(v)\,|||\,w|||,\tag{1.16}$$

where $\overline{M}(v)$ is a computable functional (usually it is presented by a certain integral over the domain Ω or by a collection of local quantities associated with finite elements). Then,

$$\sup_{w\in V_0}\left\{-\,|||\,w\,|||^2-2\ell_v(w)\right\}\leq\sup_{w\in V_0}\left\{-\,|||\,w\,|||^2+2\overline{M}(v)\,|||\,w|||\right\}=\overline{M}(v).$$

Thus,

$$|||e|||^2\leq\,\overline{M}(v)\tag{1.17}$$

and we have a guaranteed upper bound of the error. It may happen that this bound is rather coarse. Then, the integrand of $\overline{M}(v)$ does not present a good error indicator (in the sense of Definition 1.1). For example, in residual type estimates $\overline{M}(v) = C\eta(v)$, where $\eta(v)$ is a computable quantity (which is defined element wise) and C is an unknown (or known but highly overestimated constant). On the other hand, in the sense of Definition 1.3, the quantity η may be acceptable because

$$\mathbb{M}(\mathbf{m}(e)) \approx \mathbb{M}(\eta(v)). \tag{1.18}$$

We note that only this method leads to guaranteed error bounds and fully reliable error indicators.

2. Another method is to replace ℓ_v in (1.15) by a close functional, which leads to a directly computable estimator, i.e., instead of (1.16) we use

$$\ell_v(w) \approx G(v)(v)\, |||\, w||| \tag{1.19}$$

and the corresponding relation (which follows from (1.15))

$$|||e|||^2 \approx G(v). \tag{1.20}$$

This way is typical for error indicators based on post processing. The most used version is known as the *gradient averaging* indicator. Efficiency of this indicator can be justified provided that approximations possess some sort of superconvergence.

3. Another alternative is to solve the variational problem in the right-hand side of (1.15) numerically. In this case, V_0 is replaced by a sufficiently reach finite dimensional subspace V_{0h}. In fact, this leads to a version of the well-known Runge method. The most efficient versions of it lead to hierarchically based error indicators.

Below we compare several error indicators with respect to Definitions 1.1 and 1.3.

1.4 Accuracy of Error Indicators

Consider the problem

$$-\operatorname{div}(A\nabla u) = f \quad \text{in } \Omega \subset \mathbb{R}^2, \tag{1.21}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{1.22}$$

where $f = 1$ and the coefficients are strongly discontinuous, namely,

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix} \quad \text{in } \Omega_1 \quad \text{and} \quad A = \begin{bmatrix} 5 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{in } \Omega_2.$$

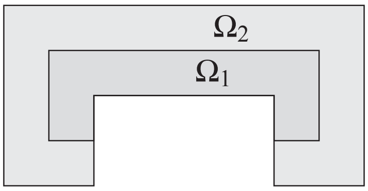


Fig. 1.2 Domain Ω

The subdomains Ω_1 and Ω_2 are depicted in Fig. 1.2. The problem belongs to the aforementioned class, where $L_2(\Omega,\mathbb{R})$, $L_2(\Omega,\mathbb{R}^2)$, $H_0^1(\Omega)$, ∇, A, and div are analogs of $\mathscr{V}$, U, V_0, Λ, $\mathscr{A}$, and Λ^*, respectively.

Approximate solutions u_h were computed by the linear Courant-type elements. In order to compare errors obtained by different error indicators with the true error, we precomputed the corresponding reference solutions using the second order Courant-type elements on a very fine mesh with 196 608 elements.

For this particular problem, among typical error indicators is

$$\eta(y)_T := \int_T (A\nabla u_h - y) \cdot \left(\nabla u_h - A^{-1}y\right) \mathrm{d}x\,, \tag{1.23}$$

where y is an approximation of the flux obtained by some numerical method. Different methods generate various error indicators. We consider

- $\mathbb{E}(y_\mathrm{G})$, where y_G is obtained by a commonly used gradient-averaging procedure. $A\nabla u_h$ is a piecewise constant and we can compute the value of y_G at node x_k as follows:

  $$y_\mathrm{G} = \frac{1}{|\omega_k|} \sum_{j \in I_{\omega_k}} \frac{1}{|T_j|} (A\nabla u_h)_{|T_j}\,,$$

  where ω_k is a patch associated with the node x_k, I_{ω_k} is the set of indices of elements on the patch ω_j, and $|\omega_k|$ and $|T_j|$ are areas of ω_k and T_j, respectively.
- $\mathbb{E}(y^0_\mathrm{RT})$, where y^0_RT is obtained by edge-wise averaging of normal fluxes on patches related to edges.
- $\mathbb{E}(y^j_\mathrm{RT})$, where y^j_RT is obtained from y^0_RT by applying the iterative quasi-equilibration procedure j times (equilibration with respect to all edges is considered one equilibration), where $\|\operatorname{div} y + f\|$ is minimized on patches related to edges. These procedures are local and for linear elements (and elementwise constant A and f) can be explicitly computed.
- $\mathbb{E}(y_\mathrm{glo})$, where y_glo is obtained by global minimization of the majorant

  $$\min_{y \in \mathrm{RT}^0} \overline{M}(v,y) := \|A\nabla u_h - y\|_{A^{-1}} + \mathrm{C}\|f + \operatorname{div} u_h\|,$$

where RT^0 is a space generated by the lowest order Raviart-Thomas elements [41] on a same mesh that was used to compute the approximation u_h. C denotes the Friedrich constant of the domain Ω.

Moreover, we consider two residual type error indicators (see [47])

- $\mathbb{E}(\eta_{RF})$ is a residual type indicator, where element-wise error contribution is

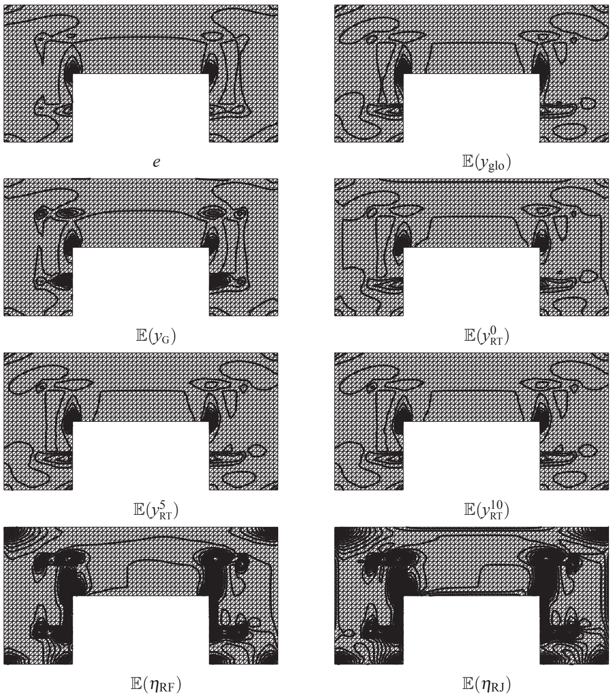
$$\eta_{RF,T} := \left(h_T^2 \|f_T\|_T^2 + \frac{1}{2} \sum_{E \in \mathscr{E}_h(T)/\mathscr{E}_{h,\partial\Omega}} |E| \, \|[n_E \cdot A \nabla u_h]_E\|_E^2 \right)^{\frac{1}{2}}, \qquad (1.24)$$

  where $E \in \mathscr{E}_h(T)/\mathscr{E}_{h,\partial\Omega}$ denotes the edges of the element T excluding the edges related to the boundary of Ω and $[\cdot]$ is the "jump" over the edge.
- $\mathbb{E}(\eta_{RJ})$ is a residual type indicator containing only jump terms,

$$\eta_{RJ,T} := \left(\frac{1}{2} \sum_{E \in \mathscr{E}_h(T)/\mathscr{E}_{h,\partial\Omega}} |E| \, \|[n_E \cdot A \nabla u_h]_E\|_E^2 \right)^{\frac{1}{2}}. \qquad (1.25)$$

In Fig. 1.3, the true error distribution and indicated element-wise error distributions are depicted for a finite element approximation computed on a regular mesh with $N = 3\,072$ elements. We see that all indicators manage to locate errors associated with corner singularities and the points where the line of discontinuity of diffusion coefficients intersects with the boundary (we note that the necessity of mesh adaptation in this area is clear a priori). However, the values of $\mathbb{E}(\eta_{RF})$ and $\mathbb{E}(\eta_{RJ})$ are substantially larger. This is also seen on histograms in Fig. 1.4, which provide another view on these results. Here, all element-wise errors are ranked in the decreasing order in accordance with the true error distribution. Thus, the very first (left) vertical bar corresponds to the element with the largest error (the number of which is 1) and the very last one to the element with the smallest error (the number of which is N). Then, the order of elements exposed along the horizontal axis is fixed and all other distributions are presented in the same order. It is clear that if $\mathbb{E}$ is accurate in the strong sense (and can be called fully reliable, see Definition 1.1), then the corresponding histogram must resemble the histogram generated by the true error. We see that not all indicators meet this condition. Similar tests have been made using finer meshes with 12 288 and 49 152 elements. They generate approximations with 7% and 4% of relative error, respectively. The corresponding histograms of the indicated errors on meshes are depicted in Figs. 1.5 and 1.6.

In Tables 1.1, 1.2, and 1.3, we measure accuracy of indicators. We use the accuracy measure in Definition 1.1. Also, the accuracy of error indicators in the sense of Definition 1.3 is evaluated with respect to three different markings: based on the average error value ($\mathbb{M}_1$), selection of 30% elements with the highest error ($\mathbb{M}_2$), and bulk criterium, where 40% of the "error mass" is selected ($\mathbb{M}_3$). Additionally, we compute the efficiency index of the majorant for computed approximations of the flux, i.e.,



e $\mathbb{E}(y_{\mathrm{glo}})$

$\mathbb{E}(y_{\mathrm{G}})$ $\mathbb{E}(y^0_{\mathrm{RT}})$

$\mathbb{E}(y^5_{\mathrm{RT}})$ $\mathbb{E}(y^{10}_{\mathrm{RT}})$

$\mathbb{E}(\eta_{\mathrm{RF}})$ $\mathbb{E}(\eta_{\mathrm{RJ}})$

Fig. 1.3 Contour lines of true and indicated error distributions for the approximation computed on a regular mesh with 3 072 elements

Table 1.1 Comparison of indicators on a regular mesh with 3 072 elements

Indicator	$\mathscr{M}(\mathbb{E})$	$\mathscr{M}(\mathbb{E},\mathbb{M}_1)$	$\mathscr{M}(\mathbb{E},\mathbb{M}_2)$	$\mathscr{M}(\mathbb{E},\mathbb{M}_3)$	I_{eff}
$\mathbb{E}(y_{\mathrm{glo}})$	0.4988	0.1204	0.0703	0.0654	1.4220
$\mathbb{E}(y_{\mathrm{G}})$	0.6877	0.1156	0.1029	0.1110	16.351
$\mathbb{E}(y^0_{\mathrm{RT}})$	0.5534	0.1243	0.0957	0.0846	24.443
$\mathbb{E}(y^5_{\mathrm{RT}})$	0.5487	0.1234	0.0755	0.0700	2.3728
$\mathbb{E}(y^{10}_{\mathrm{RT}})$	0.5643	0.1250	0.0742	0.0687	2.0144
$\mathbb{E}(\eta_{\mathrm{RF}})$	6.9200	0.2692	0.2617	0.1634	–
$\mathbb{E}(\eta_{\mathrm{RJ}})$	5.5587	0.2767	0.2617	0.1104	–

$$I_{\mathrm{eff}} := \frac{\overline{M}(u_h, y)}{\|\nabla(u - u_h)\|_A}.$$

We see that an indicator can be accurate in a weak sense with respect to a certain marker but inaccurate in the strong sense. However, in this case it might be much less accurate with respect to another marker.

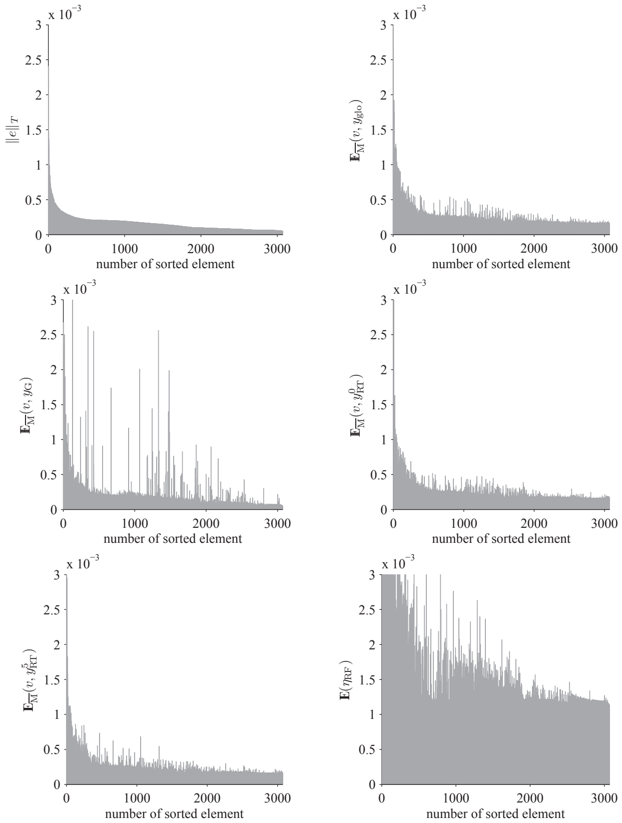


Fig. 1.4 Histograms of true and indicated error distributions for the approximation computed on a regular mesh with 3 072 elements

1.5 Accuracy of Error Indicators for Problems with Uncertain Data

Error indicators used in numerical analysis of partial differential equations usually assume that data of the problem are known exactly. In this case, a good error indicator can suggest efficient reconstructions of meshes, which lead to accurate numerical solutions. In this section, we discuss how this process may be affected by incompletely known data. Certainly this discussion is based upon rather simple examples. However, to the best of our knowledge, such type studies are quite new and our goal

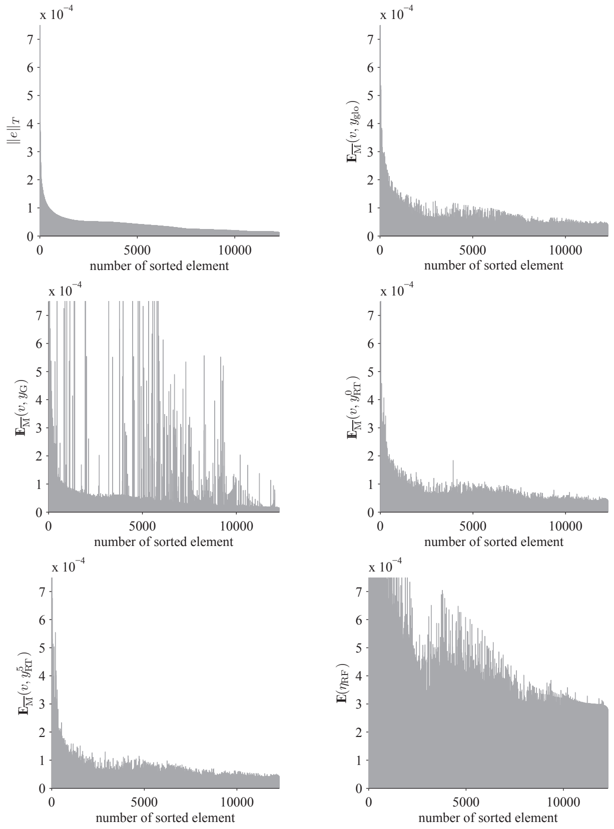


Fig. 1.5 Histograms of true and indicated error distributions for the approximation computed on a regular mesh with 12 288 elements

is to show some principal difficulties arising if error indicators are applied to problems with uncertain data. It is clear that similar difficulties will arise in many other problems.

We begin with observations motivated by Fig. 1.7 where we depict two different "error indication directions" $\mathbb{E}_1$ and $\mathbb{E}_2$. These directions are computed by means of the indicator $\mathbb{E}$ with the data D_1 and D_2, which lead to two different exact solutions u_1 and u_2, respectively. If our approximate solution v^h is far from $\mathscr{S}(\mathscr{D})$, then the directions are close (in other words if we have a coarse approximation, then good error indicators are robust with respect to small variations of data). However, this may be not true for accurate approximations. This fact does not depend on the qual-

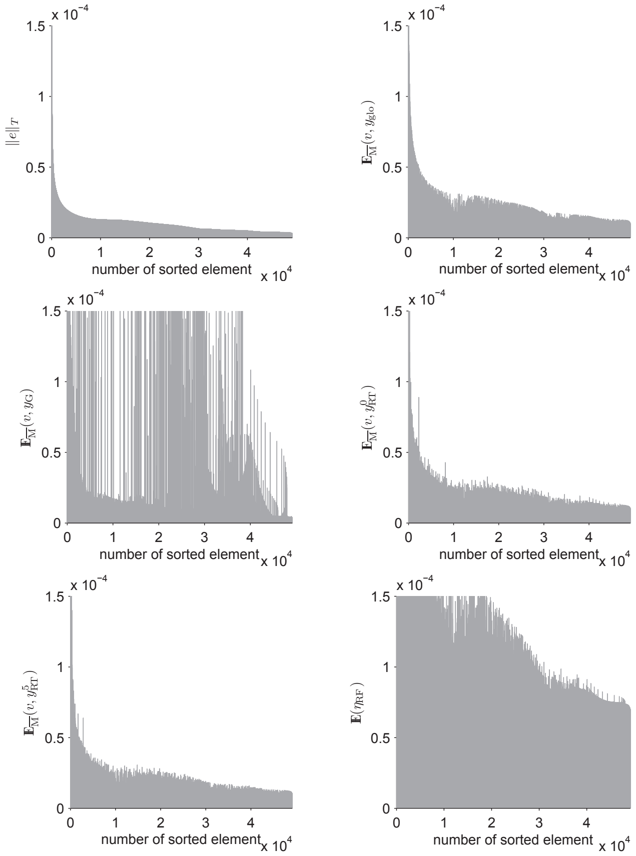


Fig. 1.6 Histograms of true and indicated error distributions for approximation computed on a regular mesh with 49 152 elements

ity of on error indicator and takes place even for the best one based on comparison of approximations and exact solutions. In practice, arrows depicted in Fig. 1.7 mean certain reconstructions of meshes. It is easy to see that if the approximate solution lies in the vicinity of $\mathscr{S}(\mathscr{D})$, then error indicators provide very different results if the data a varied within admissible bounds. Therefore, the process of sensible mesh adaptation has a limit beyond which further refinements become unreliable. Below we demonstrate this fact on a simple example. Our goal is to study how incomplete knowledge of the coefficients of diffusion coefficients impact the reliability of error indication.

Table 1.2 Comparison of indicators on a regular mesh with 12 288 elements

Indicator	$\mathscr{M}(\mathbb{E})$	$\mathscr{M}(\mathbb{E},\mathbb{M}_1)$	$\mathscr{M}(\mathbb{E},\mathbb{M}_2)$	$\mathscr{M}(\mathbb{E},\mathbb{M}_3)$	I_{eff}
$\mathbb{E}(y_{\text{glo}})$	0.4994	0.1281	0.0672	0.0545	1.4275
$\mathbb{E}(y_{\text{G}})$	1.0027	0.1192	0.0685	0.0987	32.556
$\mathbb{E}(y^0_{\text{RT}})$	0.5617	0.1245	0.0788	0.0692	48.364
$\mathbb{E}(y^5_{\text{RT}})$	0.5650	0.1303	0.0675	0.0601	3.4817
$\mathbb{E}(y^{10}_{\text{RT}})$	0.5833	0.1305	0.0669	0.0595	2.6653
$\mathbb{E}(\eta_{\text{RF}})$	6.9584	0.2636	0.2614	0.1515	–
$\mathbb{E}(\eta_{\text{RJ}})$	5.8981	0.2719	0.2614	0.0977	–

Table 1.3 Comparison of indicators on a regular mesh with 49 152 elements

Indicator	$\mathscr{M}(\mathbb{E})$	$\mathscr{M}(\mathbb{E},\mathbb{M}_1)$	$\mathscr{M}(\mathbb{E},\mathbb{M}_2)$	$\mathscr{M}(\mathbb{E},\mathbb{M}_3)$	I_{eff}
$\mathbb{E}(y_{\text{glo}})$	0.5208	0.1313	0.0653	0.0525	1.4501
$\mathbb{E}(y_{\text{G}})$	1.3685	0.1337	0.0406	0.1000	68.656
$\mathbb{E}(y^0_{\text{RT}})$	0.5807	0.1251	0.0671	0.0610	102.01
$\mathbb{E}(y^5_{\text{RT}})$	0.6059	0.1285	0.0622	0.0550	5.9855
$\mathbb{E}(y^{10}_{\text{RT}})$	0.6280	0.1295	0.0620	0.0544	4.1468
$\mathbb{E}(\eta_{\text{RF}})$	7.0463	0.2581	0.2623	0.1465	–
$\mathbb{E}(\eta_{\text{RJ}})$	6.2373	0.2665	0.2623	0.0925	–

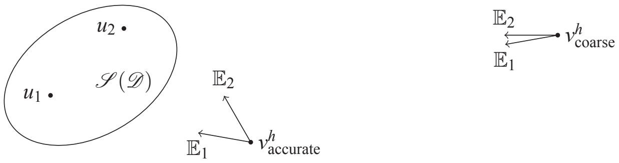


Fig. 1.7 Error indications $\mathbb{E}_1$ and $\mathbb{E}_2$ oriented towards two different solutions u_1 and u_2 in the solution set $\mathscr{S}(\mathscr{D})$

1.5.1 Numerical Experiments

In our numerical experiments, we again consider the stationary diffusion equation $\operatorname{div} A\nabla u + f = 0$ with small disturbances of the diffusion matrix $A = A_\circ + \delta B$, where the magnitude of variations δ satisfies $A_\circ\xi \cdot \xi \geq \underline{c} > \delta$, for all $|\xi| = 1$. For each element $T \in \mathscr{T}_h$, the matrix B (which defines disturbances) is symmetric and its coefficients may attain one of three values: $\{-1, 0, 1\}$. A perturbation generated in this way is clearly an extreme one. It suits our purposes, since we are trying to find perturbations generating the worst case situation which may occur with different diffusion matrices A that belong to the set D.

We note that since the amount of matrices contained in D is much larger than those representable in such a form, the sensitivity of error indicators with respect to data uncertainty is even higher than it was detected in our experiments.

Let $\mathbb{E}$ denote an error indicator computed on the set of elements $\mathscr{T}_h$ for an approximation u_h. The output of $\mathbb{E}$ is a vector $\{\mathbb{E}(u_h)\}$ that contains approximate errors value for each element in T. In computational practice, error indicators are used together with a marker $\mathbb{M}$. In this series of numerical experiments, we confine ourselves to the marker $\mathbb{M}$, which marks a certain predefined amount of elements with highest errors (denoted by N_{ref}).

Our analysis of effects caused by data uncertainty is based on the following strategy. We select a mesh $\mathcal{T}_h$ and a certain amount of matrices $A_j = A_\circ + \delta B_j$ for some given δ. For each set of data associated with the exact solution $u_j = \S(A_j)$, we compute the corresponding approximations u_{jh} on the mesh $\mathcal{T}_h$. Then, for each u_{jh}, we calculate the error indicator $\mathbb{E}_j = \mathbb{E}(A_j, u_{jh})$ and the corresponding markings $\mathbb{M}(\mathbb{E}_j)$.

The difference of two markings is natural to evaluate by means of the boolean measure analogous to that we used in (1.4). We define the quantity

$$\text{diff}(\mathbb{M}, \mathbb{E}_i, \mathbb{E}_j) := 1 - \frac{[\![\mathbb{M}(\mathbb{E}_i) \equiv \mathbb{M}(\mathbb{E}_j)]\!]}{N} \in [0,1]. \tag{1.26}$$

The quantity

$$\Theta := \max_{i,j}\{\text{diff}(\mathbb{M}, \mathbb{E}_i, \mathbb{E}_j)\} \tag{1.27}$$

shows the maximal difference produced by an error indicator with different diffusion matrices from the set D. We have tested the following commonly used error indicators.

We test the error indicators applied in the previous example, i.e., $\mathbb{E}(y_{\text{G}})$, $\mathbb{E}(y^j_{\text{RT}})$, $\mathbb{E}(y_{\text{glo}})$, $\mathbb{E}(\eta_{RF})$ and $\mathbb{E}(\eta_{RJ})$, respectively. Additionally, we introduce $\mathbb{E}(y_{\text{Gglo}})$ generated by substituting "globally averaged" y_{Gglo} in (1.23). It is calculated by globally minimizing $\|y_{\text{Gglo}} - A\nabla u_h\|^2_{A^{-1}}$ (see, e.g., [18, 11]) using the Raviart-Thomas elements.

Approximate solutions of the model problem have been computed by using the standard Courant-type finite element approximations. Indicators $\mathbb{E}(y_{\text{glo}})$ and $\mathbb{E}(y_{\text{Gglo}})$ were computed with the help of the linear Raviart-Thomas finite elements. All the problems were solved on same regular meshes, and arising systems of linear simultaneous equations were exactly solved by direct methods. In view of this fact, approximate solutions possess the Galerkin orthogonality property, and, therefore, the residual error indicator $\mathbb{E}(\eta_{RF})$ can be used. For the edge averaging indicator $\mathbb{E}(u_h, y^j_{\text{RT}})$, we set $j = 5$ (the amount of times the quasi-equilibration cycle P_{RM} is applied).

N_{elem} denotes the overall amount of elements. The marker $\mathbb{M}$ used selects 30% of elements to be refined, i.e., $N_{ref} = 0.3N_{elem}$. Note that the maximal value of Θ for this marker is 0.6. Even if markings generated by two different indicators select completely different elements, then for 40% of all elements the marked value coincides (it is zero).

We studied how the magnitude of variations δ affects error indicators and discuss typical results with the example of a simple problem where $\Omega = (0,1)^2$, $A_\circ = I$, and $f = 2(x_1(1-x_1) + x_2(1-x_2))$. The exact solution of this "mean" problem is $u_\circ = x_1(1-x_1)x_2(1-x_2)$.

The results are exposed in Table 1.4 and Fig. 1.8. They show the performance

Table 1.4 The values of Θ

(a) $\mathbb{E}(v, y_{\mathrm{G}})$, patch-wise averaging

N_{elem}	δ					
	0.005	0.01	0.02	0.03	0.04	0.05
800	0.09	0.16	0.31	0.40	0.48	0.51
3200	0.18	0.31	0.47	0.53	0.52	0.58
12800	0.32	0.48	0.52	0.59	0.60	0.60
51200	0.48	0.52	0.60	0.60	0.60	0.60
115200	0.53	0.59	0.60	0.60	0.60	0.60

(b) $\mathbb{E}(v, y^0_{\mathrm{RT}})$ edge averaging

δ					
0.005	0.01	0.02	0.03	0.04	0.05
0.09	0.16	0.30	0.40	0.45	0.50
0.16	0.30	0.46	0.52	0.53	0.52
0.30	0.46	0.53	0.56	0.59	0.59
0.46	0.53	0.59	0.60	0.60	0.60
0.52	0.57	0.59	0.60	0.60	0.60

(c) $\mathbb{E}(\eta_{RF})$ residual, full

N_{elem}	δ					
	0.005	0.01	0.02	0.03	0.04	0.05
800	0.16	0.24	0.39	0.48	0.54	0.56
3200	0.25	0.38	0.53	0.57	0.57	0.57
12800	0.38	0.53	0.57	0.58	0.59	0.59
51200	0.53	0.57	0.59	0.60	0.60	0.60
115200	0.56	0.58	0.60	0.60	0.60	0.60

(d) $\mathbb{E}(\eta_{RJ})$ residual, jumps

δ					
0.005	0.01	0.02	0.03	0.04	0.05
0.09	0.15	0.30	0.40	0.44	0.52
0.16	0.30	0.46	0.53	0.54	0.53
0.30	0.45	0.54	0.53	0.59	0.59
0.45	0.54	0.59	0.60	0.60	0.60
0.53	0.53	0.60	0.60	0.60	0.60

(e) $\mathbb{E}(v, y_{\mathrm{Gglo}})$, global averaging

N_{elem}	δ					
	0.005	0.01	0.02	0.03	0.04	0.05
800	0.08	0.15	0.30	0.39	0.46	0.50
3200	0.17	0.30	0.46	0.53	0.54	0.52
12800	0.30	0.46	0.54	0.57	0.60	0.60
51200	0.46	0.53	0.59	0.60	0.60	0.60
115200	0.52	0.57	0.60	0.60	0.60	0.60

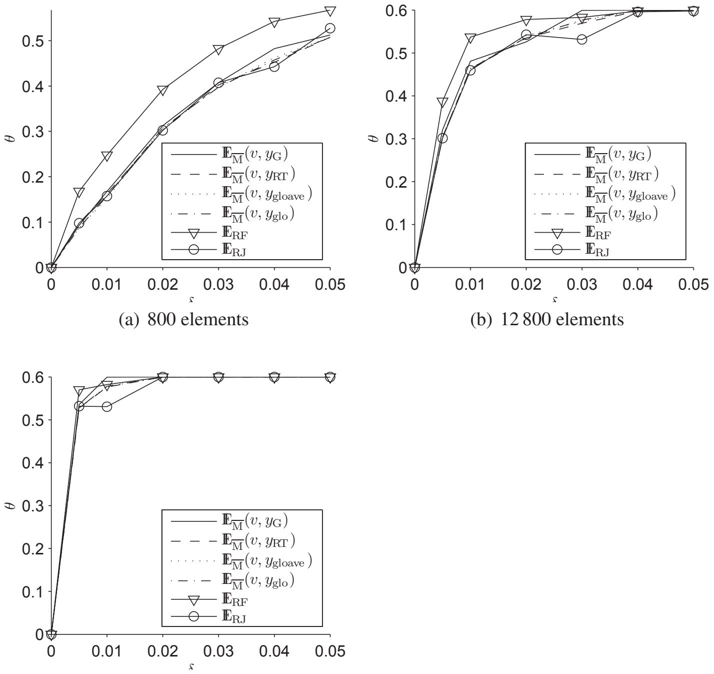
(f) $\mathbb{E}(v, y_{\mathrm{glo}})$, majorant min

δ					
0.005	0.01	0.02	0.03	0.04	0.05
0.08	0.15	0.30	0.39	0.45	0.50
0.16	0.30	0.46	0.53	0.53	0.52
0.30	0.46	0.53	0.57	0.60	0.60
0.46	0.53	0.60	0.60	0.60	0.60
0.52	0.57	0.60	0.60	0.60	0.60

of indicators on six different meshes. It is worth outlining that the actual sensitivity of error indicators with respect to the data uncertainty is even higher than in these results, because we do not consider all problems with admissible data.

Table 1.4 shows how the values of Θ (associated with the indicators) depend on the amount of elements N_{elem} and the parameter δ. It is easy to see that sufficiently small values of Θ (which correspond to relatively stable performance of an error indicator) are obtained only for small δ (such as 0.005 or 0.01) and a rather moderate amount of elements. If the values of δ are not very small (e.g., 0.04), then all the indicators may generate quite different markings. We recall that $\Theta = 0.6$ if the indicators computed for different elements of the solution set $\mathscr{D}$ may generate completely opposite markings. Obviously, this situation arises if the corresponding approximate solution lies inside (or is very close) the set $\mathscr{S}(\mathscr{D})$.

Curves in Fig. 1.8 represent these results graphically. We see that for $\delta > 0.01$ all indicators lose the reliability. *We observe that if the indeterminacy is significant compared with the approximation error, uncertainties in the matrix entries may seriously corrupt the process of error indication. This phenomenon does not depend on a particular error indicator.*



(a) 800 elements

(b) 12 800 elements

(c) 115 200 elements

Fig. 1.8 Values of Θ for different δ for three meshes

Finally, we note that in this simple test problem the effect of indicator deterioration is easy to discover even for relatively coarse meshes. However, upon our experience, similar effects will eventually arise in all problems if more and more refined meshes are used. In other words, indeterminacy of data limits efficiency (and applicability) of error indicators.

1.6 Summary and Conclusions

We presented a classification for the error indication methods and defined unified methodology to measures to evaluate and compare performance of error indicators. The application of these measures to was presented in a numerical example, where a group of established error indicators were compared. Moreover, we extended these measures to study the effects of uncertain data to error indication reliability. It was shown by a simple numerical experiment that the incomplete knowledge of the data has serious implications to the error indication, if the approximate solution is close to the accuracy limit.

References

1. Agouzal, A.: On the saturation assumption and hierarchical a posteriori error estimator. Comput. Methods Appl. Math. **2**(2), 125–131 (2002)
2. Ainsworth, M., Oden, J.T.: A procedure for a posteriori error estimation for *h*-*p* finite element methods. Comput. Methods Appl. Mech. Engrg. **101**(1–3), 73–96 (1992). Reliability in computational mechanics (Kraków, 1991)
3. Ainsworth, M., Oden, J.T.: A posteriori error estimation in finite element analysis. John Wiley & Sons, New York (2000)
4. Babuška, I., Griebel, M., Pitkäranta, J.: The problem of selecting the shape functions for a *p*-type finite element. Internat. J. Numer. Methods Engrg. **28**(8), 1891–1908 (1989)
5. Babuška, I., Rheinboldt, W.C.: A-posteriori error estimates for the finite element method. Internat. J. Numer. Meth. Engrg. **12**(10), 1597–1615 (1978)
6. Babuška, I., Rheinboldt, W.C.: Error estimates for adaptive finite element computations. SIAM J. Numer. Anal. **15**(4), 736–754 (1978)
7. Babuška, I., Rodríguez, R.: The problem of the selection of an a posteriori error indicator based on smoothening techniques. Internat. J. Numer. Methods Engrg. **36**(4), 539–567 (1993)
8. Babuška, I., Strouboulis, T.: The finite element method and its reliability. Numerical Mathematics and Scientific Computation. The Clarendon Press, Oxford University Press, New York (2001)
9. Babuška, I., Whiteman, J.R., Strouboulis, T.: Finite elements: An introduction to the method and error estimation. Oxford University Press, Oxford (2011)
10. Bangerth, W., Rannacher, R.: Adaptive finite element methods for differential equations. Birkhäuser, Basel (2003)
11. Bartels, S., Carstensen, C.: Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. II: Higher order FEM. Math. Comp. **71**(239), 971–994 (2002)
12. Bartels, S., Schreier, P.: Local coarsening of simplicial finite element meshes generated by bisections. BIT **52**(3), 559–569 (2012)
13. Becker, R., Rannacher, R.: A feed-back approach to error control in finite element methods: Basic analysis and examples. East-West J. Numer. Math. **4**(4), 237–264 (1996)
14. Besier, M., Rannacher, R.: Goal-oriented space-time adaptivity in the finite element Galerkin method for the computation of nonstationary incompressible flow. Internat. J. Numer. Methods Fluids **70**(9), 1139–1166 (2012)
15. Bonito, A., Nochetto, R.H., Pauletti, M.S.: Geometrically consistent mesh modification. SIAM J. Numer. Anal. **48**(5), 1877–1899 (2010)
16. Braess, D.: Finite elements: Theory, fast solvers, and applications in solid mechanics. Cambridge University Press, Cambridge (1997)
17. Carstensen, C.: Quasi-interpolation and a posteriori error analysis in finite element methods. M2AN Math. Model. Numer. Anal. **33**(6), 1187–1202 (1999)
18. Carstensen, C., Bartels, S.: Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. I: Low order conforming, nonconforming, and mixed FEM. Math. Comp. **71**(239), 945–969 (2002)
19. Carstensen, C., Funken, S.A.: Fully reliable localized error control in the FEM. SIAM J. Sci. Comput. **21**(4), 1465–1484 (1999)
20. Carstensen, C., Verfürth, R.: Edge residuals dominate a posteriori error estimates for low order finite element methods. SIAM J. Numer. Anal. **36**(5), 1571–1587 (1999)

21. Demkowicz, L.: Computing with hp-adaptive finite elements, Vol. 1: One and two dimensional elliptic and Maxwell problems. Chapman and Hall/CRC (2006)
22. Deuflhard, P., Leinen, P., Yserentant, H.: Concepts of an adaptive hierarchical finite element code. Impact Comput. Sci. Engrg. **1**(1), 3–35 (1989)
23. Dörfler, W., Nochetto, R.H.: Small data oscillation implies the saturation assumption. Numer. Math. **91**(1), 1–12 (2002)
24. Dörfler, W., Rumpf, M.: An adaptive strategy for elliptic problems including a posteriori controlled boundary approximation. Math. Comp. **67**(224), 1361–1382 (1998)
25. Durán, R., Muschietti, M.A., Rodriguez, R.: On the asymptotic exactness of error estimators for linear triangular finite elements. Numer. Math. **59**(1), 107–127 (1991)
26. Eriksson, K., Johnson, C.: An adaptive finite element method for linear elliptic problems. Math. Comp. **50**(182), 361–383 (1988)
27. Heimsund, B.O., Tai, X.C., Wang, J.: Superconvergence for the gradient of finite element approximations by L^2 projections. SIAM J. Numer. Anal. **40**(4), 1263–1280 (2002)
28. Houston, P., Rannacher, R., Süli, E.: A posteriori error analysis for stabilised finite element approximations of transport problems. Comput. Methods Appl. Mech. Engrg. **190**(11–12), 1483–1508 (2000)
29. Johnson, C., Hansbo, P.: Adaptive finite elements in computational mechanics. Comput. Methods Appl. Mech. Engrg. **101**(1–3), 143–181 (1992)
30. Křížek, M., Neittaanmäki, P.: Superconvergence phenomenon in the finite element method arising from averaging gradients. Numer. Math. **45**(1), 105–116 (1984)
31. Kuzmin, D., Möller, M.: Goal-oriented mesh adaptation for flux-limited approximations to steady hyperbolic problems. J. Comput. Appl. Math. **233**(12), 3113–3120 (2010)
32. Křížek, M., Neittaanmäki, P., Stenberg, R. (eds.): Finite element methods: Superconvergence, postprocessing and a posteriori estimates (Jyväskylä, 1996), *Lecture Notes in Pure and Appl. Math.*, vol. 196. Marcel Dekker, New York (1998)
33. Ladevéze, P., Leguillon, D.: Error estimate procedure in the finite element method and applications. SIAM J. Numer. Anal. **20**(3), 485–509 (1983)
34. Mali, O., Repin, S., Neittaanmäki, P.: Accuracy verification methods, theory and algorithms. Springer, Berlin. In print
35. Mommer, M.S., Stevenson, R.: A goal-oriented adaptive finite element method with convergence rates. SIAM J. Numer. Anal. **47**(2), 861–886 (2009)
36. Oden, J.T., Prudhomme, S.: Goal-oriented error estimation and adaptivity for the finite element method. Comput. Math. Appl. **41**(5–6), 735–756 (2001)
37. Peraire, J., Patera, A.T.: Bounds for linear-functional outputs of coercive partial differential equations: Local indicators and adaptive refinement. In: P. Ladevéze, J.T. Oden (eds.) Advances in Adaptive Computational Methods in Mechanics (Cachan, 1997), pp. 199–216. Elsevier, Amsterdam (1998)
38. Porta, G.M., Perotto, S., Ballio, F.: Anisotropic mesh adaptation driven by a recovery-based error estimator for shallow water flow modeling. Internat. J. Numer. Methods Fluids **70**(3), 269–299 (2012)
39. Rannacher, R.: The dual-weighted-residual method for error control and mesh adaptation in finite element methods. In: J. Whiteman (ed.) The Mathematics of Finite Elements and Applications, X, MAFELAP 1999 (Uxbridge), pp. 97–116. Elsevier, Oxford (2000)
40. Rannacher, R., Vexler, B.: Adaptive finite element discretization in PDE-based optimization. GAMM-Mitt. **33**(2), 177–193 (2010)
41. Raviart, P.A., Thomas, J.M.: Primal hybrid finite element methods for 2nd order elliptic equations. Math. Comp. **31**(138), 391–413 (1977)
42. Rheinboldt, W.C.: On a theory of mesh-refinement processes. SIAM J. Numer. Anal. **17**(6), 766–778 (1980)
43. Shepherd, J.F., Dewey, M.W., Woodbury, A.C., Benzley, S.E., Staten, M.L., Owen, S.J.: Adaptive mesh coarsening for quadrilateral and hexahedral meshes. Finite Elem. Anal. Des. **46**(1–2), 17–32 (2010)
44. Sirois, Y., McKenty, F., Gravel, L., Guibault, F.: Hybrid mesh adaptation applied to industrial numerical combustion. Internat. J. Numer. Methods Fluids **70**(2), 222–245 (2012)

45. Stein, E., Ohnimus, S.: Coupled model- and solution-adaptivity in the finite element method. Comput. Methods Appl. Mech. Engrg. **150**(1–4), 327–350 (1997)
46. Stein, E., Rüter, M., Ohnimus, S.: Error-controlled adaptive goal-oriented modeling and finite element approximations in elasticity. Comput. Methods Appl. Mech. Engrg. **196**(37–40), 3598–3613 (2007)
47. Verfürth, R.: A review of a posteriori error estimation and adaptive mesh-refinement techniques. Wiley–Teubner, New York (1996)
48. Wahlbin, L.B.: Superconvergence in Galerkin finite element methods, *Lecture Notes in Mathematics*, vol. 1605. Springer, Berlin (1995)
49. Wang, J.: Superconvergence analysis for finite element solutions by the least-squares surface fitting on irregular meshes for smooth problems. J. Math. Study **33**(3), 229—243 (2000)
50. Wang, X., Ye, X.: Superconvergence analysis for the Navier-Stokes equations. Appl. Numer. Math. **41**(4), 515–527 (2002)
51. Zhang, Z., Naga, A.: A new finite element gradient recovery method: Superconvergence property. SIAM J. Sci. Comput. **26**(4), 1192–1213 (2005)
52. Zhu, J.Z., Zienkiewicz, O.C.: Adaptive techniques in the finite element method. Comm. Appl. Numer. Methods **4**(2), 197–204 (1988)
53. Zienkiewicz, O.C., Boroomand, B., Zhu, J.Z.: Recovery procedures in error estimation and adaptivity: Adaptivity in linear problems. In: P. Ladeveze, J.T. Oden (eds.) Advances in Adaptive Computational Methods in Mechanics (Cachan, 1997), *Stud. Appl. Mech.*, vol. 47, pp. 3–23. Elsevier, Amsterdam (1998)
54. Zienkiewicz, O.C., Zhu, J.Z.: A simple error estimator and adaptive procedure for practical engineering analysis. Internat. J. Numer. Meth. Engrg. **24**(2), 337–357 (1987)