

Inaugural-Dissertation
zur
Erlangung der Doktorwürde
der
Naturwissenschaftlich-Mathematischen Gesamtfakultät
der
Ruprecht-Karls-Universität
Heidelberg

vorgelegt von
Diplom Mathematiker Boris Vexler
aus Moskau

Tag der mündlichen Prüfung: 3. Mai 2004

Adaptive Finite Element Methods for Parameter Identification Problems

January 7, 2004

Gutachter: Prof. Dr. Rolf Rannacher
Prof. Dr. Hans-Georg Bock

Contents

Chapter 1. Introduction	7
Chapter 2. Theoretical Results	11
2.1. Formulation of the problem	11
2.2. Existence of a solution	13
2.3. Necessary and sufficient optimality conditions	21
2.4. Stability and uniqueness of the solution	24
2.5. Statistical considerations	29
Chapter 3. Finite Element Discretization	33
3.1. Triangulations and finite element spaces	33
3.2. Discretization of a parameter identification problem	35
3.3. Existence of a solution of the discrete problem	37
3.4. A priori error analysis	40
3.5. Numerical examples	47
Chapter 4. Optimization Algorithm	49
4.1. Newton type methods	49
4.2. Realizations of the algorithm	51
4.3. Trust region method	54
4.4. Numerical examples	57
Chapter 5. A Posteriori Error Analysis	61
5.1. An introductory example	63
5.2. A posteriori error estimation for the error in parameters	65
5.3. An extension to more general error functionals	72
5.4. Localization of the error estimator	74
5.5. Numerical examples	76
Chapter 6. Application to CFD Problems	87
6.1. Discretization of the Navier-Stokes equations	87
6.2. Treatment of parameter-depended boundary conditions	89
6.3. Bypass simulation with unknown boundary conditions	90
6.4. Flow in a canal with a rough wall	93

Chapter 7. Application to Multidimensional Reactive Flows	97
7.1. Identification of Arrhenius parameters	97
7.2. Identification of diffusion parameters	101
Chapter 8. Conclusion and Future Work	107
Bibliography	109

CHAPTER 1

Introduction

Usually in Mathematics you have an equation and you want to find a solution. Here you are given a solution and you have to find the equation. I like that.

Julia Robinson

In this thesis we analyze parameter identification problems governed by partial differential equations and develop efficient numerical methods for their solution.

Often, a physical (chemical) model described by a system of partial differential equations involves unknown parameters, which cannot be measured directly, or whose measurement would require too much effort. For instance, this can appear by modeling of heuristic laws (like the Arrhenius law), by calibration of transport models or by covering unknown boundary conditions. In all these situations, the estimation (identification) of the unknown parameters is indispensable for successful simulation and optimization of the corresponding physical processes. The information required for parameter identification is usually obtained by observations of measurable quantities, like forces, fluxes, point values of pressure, velocity or concentration. Thereafter, the unknown parameters are determined such that the discrepancy between the measured quantities and the corresponding quantities obtained by solving the underlying system of partial differential equations is minimal. Choosing a way of measuring this discrepancy (in an appropriate norm) one obtains an optimization (minimization) problem to be solved.

The aim of this work is the analysis of parameter identification problems and the development of efficient numerical algorithms for their solution, based on adaptive finite element methods. Since, in general, the computational effort for solving the arising optimization problems exceeds significantly the cost for a simple simulation, the question of choosing efficient (cheap) discretizations is crucial for applications. Our approach to this question is based on the a posteriori error estimation for finite element discretization of the problem. We derive a posteriori error estimators to be used in an adaptive mesh refinement algorithm producing economical meshes for parameter identification.

The concepts of adaptivity based on a posteriori error estimation are now commonly accepted for numerical solution of partial differential equation. In this work we provide the first

systematic approach to adaptive finite element discretization of parameter identification problems. This is based on the works on a posteriori error estimation for optimal control problems by Becker & Rannacher [18, 19] and Becker [11]. However, substantial extensions of the techniques from [11, 18, 19] were performed for covering parameter identification problems.

The methods developed in this thesis are applied to parameter identification in fluid dynamics and to estimation of chemical models in multidimensional reactive flow problems. In the applications under consideration there is a finite number of the unknown parameters, i.e. the parameters build a vector from a finite dimensional space. In this work we consider only the case of finite dimensional parameter spaces, in contrast to the estimation of so called *distributed* parameters. This is due to the fact, that the analysis and the required solution algorithms for parameter identification problems differ depending on the dimension (finite or infinite) of the parameter space. However, most of the results presented below can be extended to the case of distributed parameters, which is a subject of future work.

The organization of the thesis is as follows: In the next chapter we discuss the formulation of parameter identification problems as an optimization problem. Thereafter we characterize possible solutions by necessary and sufficient optimality conditions and analyze existence, uniqueness and stability of them. Moreover, we discuss some statistical aspects of parameter identification, in particular the question of the statistical quality of the estimated parameters.

In Chapter 3 we treat the finite element discretization of parameter identification problems. The main purpose of this chapter is the development of a priori error analysis for the error in parameters due to the discretization. We derive a priori error estimates for parameter identification problems governed by elliptic partial differential equations of second order and discuss possible extensions of our approach.

The solution algorithm for the discretized problem is discussed in Chapter 4. We describe different Newton type methods applied to the unconstrained (reduced) formulation of the parameter identification problem. Moreover, we discuss trust region techniques for globalization of convergence. The behavior of the algorithms is demonstrated for an example problem.

Chapter 5 is devoted to a posteriori error analysis. Here, we develop an a posteriori error estimator for the error in parameters due to the discretization. Its purpose is to guide an adaptive mesh refinement algorithm producing a sequence of economical, locally refined meshes. Furthermore, it is used to assess the accuracy of the computed parameters. Exploiting the special structure of the parameter identification problem, allows us to derive an error estimator which is cheap in comparison to the overall optimization algorithm. Several examples illustrate the behavior of the adaptive mesh refinement algorithm based on our error estimator.

In Chapter 6 the presented methods are applied to parameter identification problems governed by the incompressible Navier-Stokes equations. We consider two problems, where the exact boundary conditions are unknown. This describes a typical difficulty in computational fluid dynamics. We formulate these problems as parameter identification problems with parameterized boundary conditions and treat them by the techniques from previous chapters. The numerical results show the capability of our methods.

An application to parameter estimation in multidimensional reactive flow problems is discussed in Chapter 7. We consider two typical problems: estimation of the Arrhenius parameters for a simple combustion model and calibration of the diffusion coefficients for a hydrogen flame with detailed chemistry. Here, the underlying model includes the compressible Navier-Stokes equations and nine (nonlinear) convection-diffusion-reaction equations for chemical species. To the author's knowledge, this is the first published result on automatic parameter estimation for multidimensional computation of flames.

In the last chapter, conclusions and an outlook on future work are given. Here, we summarize the results presented in this thesis and discuss some extension ideas.

ACKNOWLEDGMENTS

I would like to express my gratitude to Prof. Dr. Rolf Rannacher and Prof. Dr. Roland Becker for suggesting this interesting subject and for continuous supporting this work. Furthermore, I would like to thank Dr. Malte Braack and Dipl. Math. Thomas Richter for many fruitful discussions.

This work has been supported by the German Research Foundation (DFG), through the *Graduiertenkolleg* 'Complex Processes: Modeling, Simulation and Optimization' at the Interdisciplinary Center of Scientific Computing (IWR), University of Heidelberg and the *Sonderforschungsbereich* 359 'Reactive Flows, Diffusion and Transport'.

CHAPTER 2

Theoretical Results

In this chapter we describe considered parameter identification problems from the theoretical point of view. In Section 2.1 we start with the formulation of the parameter identification problem and introduce some notation to be used throughout. Section 2.2 is devoted to the question of existence of a solution for problems under consideration. Thereafter, in Section 2.3 the solutions are characterized by necessary and sufficient optimality condition. In Section 2.4 we introduce the concept of a stable solution and analyze the main properties of it. In the last section of this chapter we touch some statistical aspects of parameter identification.

2.1. Formulation of the problem

We consider parameter identification problems involving a finite number of unknown parameters in the following abstract form: The state variable u in an appropriate Hilbert space V is determined by a partial differential equation (*state equation*) in weak form:

$$a(u, q)(\phi) = f(\phi) \quad \forall \phi \in V, \quad (2.1.1)$$

where $q \in Q = \mathbb{R}^{n_p}$ denotes the unknown parameters. The semi-linear form $a(\cdot, \cdot)(\cdot)$ is defined on the Hilbert space $V \times Q \times V$. Semi-linear forms are written with two parentheses, the first one refers to the nonlinear arguments, whereas the second one embraces all linear arguments. The partial derivatives of the semi-linear form $a(\cdot, \cdot)(\cdot)$ are denoted by $a'_u(\cdot, \cdot)(\cdot, \cdot)$, $a'_q(\cdot, \cdot)(\cdot, \cdot)$ etc. The linear functional $f \in V'$ represents the right hand side of the state equation, where V' denotes the dual space of V .

REMARK 2.1.1. For the treatment of Dirichlet boundary conditions the state variable u is searched in an affine space $\hat{u} + V$, where $\hat{u} \in \hat{V}$ describes the boundary conditions in an appropriate Hilbert space $\hat{V} \supset V$. In the following we assume $\hat{V} = V$ and $\hat{u} = 0$ for simplicity of notation.

Further, we are given an observation operator $C : V \rightarrow Z$, which maps the *state variable* u to the space of measurements $Z = \mathbb{R}^{n_m}$, where we assume that $n_m \geq n_p$. We denote by $\langle \cdot, \cdot \rangle_Z$ a scalar product of Z and by $\| \cdot \|_Z$ the corresponding norm. Similar notation are used for the scalar product and norm in the space Q .

The values of the parameters are estimated from a given set of measurements $\bar{C} \in Z$ using a least squares approach such that we obtain a constrained optimization problem with the cost

functional $J : V \times Q \rightarrow \mathbb{R}$:

$$\text{Minimize } J(u, q) := \frac{1}{2} \|C(u) - \bar{C}\|_Z^2 + \frac{\alpha}{2} \|q - \bar{q}\|_Q^2, \quad (2.1.2)$$

under the constraint (2.1.1). Here, the cost functional (2.1.2) is the sum of the squared norm of the so called *least squares residual* defined by

$$R^{LS}(u) := \bar{C} - C(u), \quad (2.1.3)$$

and a regularization term involving prescribed $\alpha \geq 0$ and $\bar{q} \in Q$.

For covering additional constraints on the parameters we seek $q \in Q_{ad}$, where $Q_{ad} \subset Q$ is an admissible set of parameters. A pair $(u, q) \in V \times Q$ is called admissible, if $q \in Q_{ad}$ and (u, q) fulfills the state equation (2.1.1). An admissible pair (u, q) is a *local* solution of the problem (2.1.1, 2.1.2), if there exist neighborhoods $B_u \subset V$ of u and $B_q \subset Q$ of q such that:

$$J(\tilde{u}, \tilde{q}) \geq J(u, q), \quad (2.1.4)$$

for all admissible pairs $(\tilde{u}, \tilde{q}) \in B_u \times B_q$. If the condition (2.1.4) holds for all admissible pairs $(\tilde{u}, \tilde{q}) \in V \times Q_{ad}$, the solution (u, q) is called a *global* solution.

Below, we introduce some notations from the theory of function spaces, which we will use throughout. Properties of this spaces can be found e.g. in Alt [2], Riesz & Sz.-Nagy [82] or Zeidler [89].

NOTATION FOR FUNCTION SPACES

For a domain $\Omega \subset \mathbb{R}^d$ we denote the Lebesgue space of square-integrable functions on Ω by $L^2(\Omega)$. It is a Hilbert space with scalar product and norm:

$$(v, w)_\Omega = \int_\Omega vw \, dx, \quad \|v\|_\Omega = \left(\int_\Omega |v|^2 \, dx \right)^{\frac{1}{2}}.$$

Analogous, $L^2(\partial\Omega)$ denotes the space of square-integrable functions on the boundary $\partial\Omega$ equipped with the scalar product and norm:

$$(v, w)_{\partial\Omega} = \int_{\partial\Omega} vw \, ds, \quad \|v\|_{\partial\Omega} = \left(\int_{\partial\Omega} |v|^2 \, ds \right)^{\frac{1}{2}}.$$

The Sobolev spaces $H^1(\Omega)$ and $H^2(\Omega)$ consist of those functions $v \in L^2(\Omega)$ which possess first- and second-order (distributional) derivatives $\nabla v \in L^2(\Omega)^d$ and $\nabla^2 v \in L^2(\Omega)^{d \times d}$, respectively. For this spaces we use the norms:

$$\|v\|_{H^1(\Omega)} = \left(\|v\|_\Omega^2 + \|\nabla v\|_\Omega^2 \right)^{\frac{1}{2}}, \quad \|v\|_{H^2(\Omega)} = \left(\|v\|_\Omega^2 + \|\nabla v\|_\Omega^2 + \|\nabla^2 v\|_\Omega^2 \right)^{\frac{1}{2}}.$$

Further, there exists a continuous trace operator $\gamma : H^1(\Omega) \rightarrow L^2(\partial\Omega)$ with $v|_{\partial\Omega} := \gamma(v)$. It allows, to define the function space $H_0^1(\Omega)$ by:

$$H_0^1(\Omega) = \{v \in H^1(\Omega) \mid v|_{\partial\Omega} = 0\}.$$

By the Poincaré inequality,

$$\|v\|_{\Omega} \leq c \|\nabla v\|_{\Omega} \quad \forall v \in H_0^1(\Omega),$$

the H^1 -seminorm $\|\nabla v\|_{\Omega}$ is a norm on the subspace $H_0^1(\Omega)$. The spaces $L^p(\Omega)$ and $W^{m,p}(\Omega)$ are defined in a similar way, see e.g. Alt [2].

If Ω describe the domain on which the differential equation is posed, we usually omit the subscript Ω in the notations of norms and scalar products, for instance $\|v\| = \|v\|_{\Omega}$. All the above notation will be synonymously used also for vector- or matrix-valued functions v .

2.2. Existence of a solution

The problem (2.1.1, 2.1.2) belongs to a big class of optimal control problems involving partial differential equation. There is a number of publications, where the question of the existence of a solution of such problems is discussed, see e.g. Lions [71], Fursikov [52], Litvinov [72] and Neittaanmaki & Tiba [76]. For some classes of parameter identification problems with distributed parameters this question is analyzed e.g. in Banks & Kunisch [9] and Kravaris & Seinfeld [70].

Due to the fact, that the space of parameters in the problem under consideration has finite dimension n_p , the question of the existence of a solution can be treated in a slightly simpler way, as for the problems with distributed parameters. However, the space of measurements Z is also finite dimensional. It often leads to the fact, that the boundedness of the cost function (2.1.2) does not necessary imply the boundedness of the state variable and therefore the so called *coercivity relation*, see Fursikov [52], is usually not valid.

In this section we prove the existence of a solution of the problem (2.1.1, 2.1.2) for two typical sets of assumptions and also for two example problems, where some of the assumptions are not fulfilled.

Throughout we assume the semilinear form $a(\cdot, \cdot)(\cdot)$ and the observation operator $C(\cdot)$ to be continuously differentiable and introduce an operator $A : V \times Q \rightarrow V'$ by the following identification:

$$\langle A(u, q), \phi \rangle_{V' \times V} = a(u, q)(\phi) \quad \forall \phi \in V, \quad (2.2.5)$$

where $\langle \cdot, \cdot \rangle_{V' \times V}$ denotes the duality pairing between the Hilbert space V and its dual V' . Using this notation we can rewrite the state equation (2.1.1) in an operator form:

$$A(u, q) = f,$$

and formulate the following assumption for it:

ASSUMPTION (A1). *There exists a nonempty open subset Q_0 of Q such that the state equation (2.1.1) possesses an unique solution $u \in V$ for all $q \in Q_0$.*

In addition we make the following regularity assumptions for the partial derivative A'_u of the operator A :

ASSUMPTION (A2). *For an arbitrary $q \in Q_0$ and the solution $u \in V$ of (2.1.1) the partial derivative $A'_u(u, q)$ of the operator A is a linear homeomorphism of V onto V' .*

These two assumptions guarantee the existence of a solution operator by following proposition:

PROPOSITION 2.2.1. *Assume (A1) and (A2) are valid. Then there exists a continuously differentiable solution operator $S : Q_0 \rightarrow V$ such that there holds for all $q \in Q_0$:*

$$A(S(q), q) = f,$$

or equivalently

$$a(S(q), q)(\phi) = f(\phi) \quad \forall \phi \in V.$$

PROOF. The proof is given by a direct application of the implicit function theorem, see, e.g., Dieudonné [40]. \square

REMARK 2.2.2. The solution operator S inherits the regularity of the operator A , i.e. if the operator A is of C^k -class ($k \geq 1$) then the solution operator S is of C^k -class too.

We restrict our consideration to the case, where the state equation possesses a solution for all $q \in Q_{ad}$, i.e. we make the following assumption:

ASSUMPTION (A3). $Q_{ad} \subset Q_0$.

The existence of a solution operator S allows us to formulate the two following theorems about the existence of a solution of the problem (2.1.1, 2.1.2).

THEOREM 2.2.1. *Let Assumptions (A1), (A2) and (A3) be valid and the admissible set Q_{ad} be bounded and closed in Q . Then the problem (2.1.1, 2.1.2) possesses a solution $(u, q) \in V \times Q_{ad}$.*

PROOF. Let \mathcal{A} be a set of admissible elements, defined by

$$\mathcal{A} := \{(u, q) \in V \times Q_{ad} \mid A(u, q) = f\}. \quad (2.2.6)$$

Due to Assumptions (A1) and (A3) the set \mathcal{A} is nonempty and due to the definition of J (2.1.2) there holds:

$$J(\mathcal{A}) \subset \mathbb{R}_+ \cup \{0\}.$$

This implies the existence of a nonnegative real number J^* with

$$J^* = \inf_{(u, q) \in \mathcal{A}} J(u, q)$$

and a sequence $\{(u_n, q_n)\} \subset \mathcal{A}$ with

$$\lim_{n \rightarrow \infty} J(u_n, q_n) = J^*.$$

Due to the boundedness of the finite dimensional set Q_{ad} there exists a subsequence $\{q_{n_k}\} \subset Q_{ad}$ and $q \in Q$ with

$$\lim_{k \rightarrow \infty} q_{n_k} = q.$$

The fact, that Q_{ad} is closed, implies that $q \in Q_{ad}$.

We set $u := S(q)$ and show, that $(u, q) \in V \times Q_{ad}$ is a solution of the problem (2.1.1, 2.1.2).

Due to the continuity of the solution operator S and the observation operator C we obtain:

$$\lim_{k \rightarrow \infty} J(S(q_{n_k}), q_{n_k}) = J(u, q).$$

This implies that $J(u, q) = J^*$. This completes the proof. \square

For this result we do not need any assumption on the regularization parameter α . It means, that Theorem 2.2.1 can be applied in the case $\alpha = 0$. The next result provides the existence of a solution of the problem (2.1.1, 2.1.2) in the case, where we do not require the boundedness of Q_{ad} , but the regularization parameter α is chosen to be positive.

THEOREM 2.2.2. *Let Assumptions (A1), (A2) and (A3) be valid, the regularization parameter α be positive and the admissible set Q_{ad} be closed in Q . Then the problem (2.1.1, 2.1.2) possesses a solution $(u, q) \in V \times Q_{ad}$.*

PROOF. The proof is similar to the one of Theorem 2.2.1. Here, we use the boundedness of the sequence $\{J(u_n, q_n)\}$ and due to the fact, that $\alpha > 0$ we obtain the boundedness of the sequence $\{q_n\}$. \square

These two theorems guarantee the existence of a solution for a big class of parameter identification problems. However, there are a lot of situations, where some of the conditions, required in Theorem 2.2.1 and Theorem 2.2.2, are not satisfied. In the following we prove the existence of solutions for two examples, where the above theorems cannot be applied. These examples are prototypical for the problems we consider later on.

EXAMPLE 1

In the first example we consider a simplified version of the application problem described in Chapter 7 (Estimation of Arrhenius parameter). The state u is given by the following nonlinear elliptic equation:

$$\begin{aligned} -\Delta u + e^{qu} &= f && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{2.2.7}$$

where $\Omega \subset \mathbb{R}^2$ is a convex domain with the polygonal boundary $\partial\Omega$ and $f \in L_2(\Omega)$ is a right hand side function with

$$f(x) > 1 \quad a.e. x \in \Omega. \quad (2.2.8)$$

The parameter $q \in Q_{ad} = \mathbb{R}_+ \cup \{0\}$ is estimated using measurements given by the mean values of the state variable along fixed measurement lines Γ_i , see Figure 2.2.1.

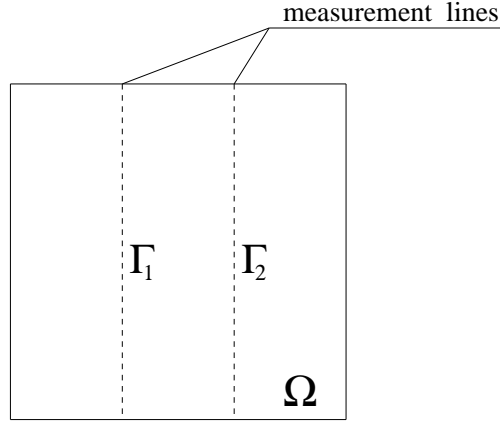


FIGURE 2.2.1. The computational domain with measurement lines.

The components of the corresponding observation operator C have the following form:

$$C_i(v) = \int_{\Gamma_i} v ds \quad i = 1 \dots n_m,$$

and the parameter identification problem is formulated as follows: For $(u, q) \in V \times Q_{ad}$ with $V = H_0^1(\Omega)$:

$$\text{Minimize } J(u) := \frac{1}{2} \sum_{i=1}^{n_m} (C_i(u) - \bar{C}_i)^2 \quad (2.2.9)$$

under the constraint (2.2.7). The components of the measurement vector $\bar{C} \in Z = \mathbb{R}^{n_m}$ are denoted by \bar{C}_i .

In this example we have neither the boundedness of the admissible set Q_{ad} nor a regularization in contrast to the class of the parameter identification problems considered in Theorem 2.2.1 and Theorem 2.2.2.

However, due to the fact, that the nonlinear differential operator A of the equation (2.2.7) is monotone (see e.g. Zeidler [89]) the equation (2.2.7) possesses a unique solution for all $q \in Q_{ad}$, i.e. Assumption (A1) is valid. Due to the ellipticity and the boundedness of the operator A Assumption (A2) is valid as well. This implies the existence of a continuously differentiable

solution operator S (see Proposition 2.2.1) and a nonnegative real number

$$J^* = \inf_{(u,q) \in \mathcal{A}} J(u), \quad (2.2.10)$$

where \mathcal{A} is the set of admissible elements, defined as in (2.2.6). A result of solvability of the problem (2.2.9) is given in the following proposition:

PROPOSITION 2.2.3. *Let for the optimal value J^* the following condition be fulfilled:*

$$J^* < J(0) = \frac{1}{2} \|\bar{C}\|^2, \quad (2.2.11)$$

then the problem (2.2.9) possesses a solution $(u, q) \in V \times Q_{ad}$.

PROOF. Analogous to the proof of Theorem 2.2.1 there exists a sequence $\{(u_n, q_n)\} \subset \mathcal{A}$ with

$$\lim_{n \rightarrow \infty} J(u_n) = J^*.$$

We rewrite the state equation (2.2.7) for $u_n \in V$ in a weak form:

$$(\nabla u_n, \nabla \phi) + (e^{q_n u_n}, \phi) = (f, \phi) \quad \forall \phi \in V,$$

set $\phi = u_n$ and obtain:

$$\|\nabla u_n\|^2 + (e^{q_n u_n}, u_n) = (f, u_n).$$

By virtue of the maximum principle (see e.g. Hellwig [62]) we have that the state variable u_n is positive due to the condition (2.2.8). Therefore, using the fact, that $e^x > 1 + x \quad \forall x \in \mathbb{R}$ we obtain:

$$\|\nabla u_n\|^2 + (1 + q_n u_n, u_n) \leq (f, u_n).$$

Then, using Cauchy-Schwarz and Young's inequalities:

$$\|\nabla u_n\|^2 + q_n \|u_n\|^2 \leq (f - 1, u_n) \leq \|f - 1\| \|u_n\| \leq \frac{1}{4q_n} \|f - 1\|^2 + q_n \|u_n\|^2$$

and consequently:

$$\|\nabla u_n\|^2 \leq \frac{1}{4q_n} \|f - 1\|^2. \quad (2.2.12)$$

If the sequence $\{q_n\}$ is unbounded, then there exists a subsequence $\{q_{n_k}\} \subset Q_{ad}$ such that $q_{n_k} \rightarrow \infty$ for $k \rightarrow \infty$ and due to (2.2.12) $u_{n_k} \rightarrow 0$ in V . By virtue of the trace theorem (see e.g. Alt [2]) we obtain that

$$C_i(u_{n_k}) = \int_{\Gamma_i} u_{n_k} ds \rightarrow 0 \quad \text{for } k \rightarrow \infty, \quad i = 1 \dots n_m$$

and consequently $J(u_{n_k}) \rightarrow \frac{1}{2} \|\bar{C}\|^2$, which is a contradiction with (2.2.11). Therefore the sequence $\{q_n\}$ is bounded and we complete the proof analogous to Theorem 2.2.1. \square

REMARK 2.2.4. The crucial assumption (2.2.11) can be simply checked for a given set of measurements \bar{C} . It is equivalent to the following condition: There exists $q_0 \in Q_{ad}$ with $J(S(q_0)) < \frac{1}{2}\|\bar{C}\|$.

EXAMPLE 2

In the second example we consider the following diffusion-reaction equation with unknown diffusion coefficient q in a convex domain Ω with the polygonal boundary $\partial\Omega$:

$$\begin{aligned} -q\Delta u + u &= 1 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{2.2.13}$$

The parameter q is estimated using measurements given by the values of the state variable at n_m different points ξ_i and the components of the corresponding observation operator C have the following form:

$$C_i(v) = v(\xi_i) \quad i = 1 \dots n_m. \tag{2.2.14}$$

The corresponding parameter identification problem is formulated as follows: For $(u, q) \in V \times Q_{ad}$ with $V = H_0^1(\Omega)$

$$\text{Minimize } J(u) := \frac{1}{2} \sum_{i=1}^{n_m} (u(\xi_i) - \bar{C}_i)^2 \tag{2.2.15}$$

under the constraint (2.2.13). The components of the measurement vector $\bar{C} \in Z = \mathbb{R}^{n_m}$ are denoted by \bar{C}_i .

The natural choice for the admissible set of parameters is $Q_{ad} = \mathbb{R}_+$, the largest connected set such that the equation (2.2.13) possesses an unique solution for all $q \in Q_{ad}$. This set is neither bounded nor closed. Moreover, the observation operator C is here an unbounded linear operator on V (see, e.g. Riesz & Sz.-Nagy [82]). This complicates the proof of the existence of a solution of the problem (2.2.15). We start with a lemma about a singularly perturbed problem, which we use later on for the proof of the solvability of the problem (2.2.15).

LEMMA 2.2.5. *Let $v \in H^1(\Omega)$ be determined by the equation*

$$\begin{aligned} -\varepsilon\Delta v + v &= 0 && \text{in } \Omega, \\ v &= 1 && \text{on } \partial\Omega, \end{aligned} \tag{2.2.16}$$

where Ω is a convex domain with the polygonal boundary $\partial\Omega$. Moreover, let for $0 < \delta < \text{diam}(\Omega)$ the domain Ω_δ be defined by:

$$\Omega_\delta := \{x \in \Omega \mid \text{dist}(x, \partial\Omega) > \delta\}.$$

Then there holds:

$$\|v\|_{\Omega_\delta} \leq 2 \text{diam}(\Omega)^{\frac{1}{2}} e^{-\frac{\delta}{2\sqrt{\varepsilon}}}.$$

PROOF. The proof is given analogous to the proof of Lemma 1 from Rannacher [80]. Let the functions $d, \sigma : \Omega \rightarrow \mathbb{R}$ be given by: $d(x) := \text{dist}(x, \partial\Omega)$ and

$$\sigma(x) := \min\{e^{\frac{d(x)}{\sqrt{\varepsilon}}}, e^{\frac{\delta}{\sqrt{\varepsilon}}}\}.$$

Then, there holds:

$$\varepsilon \|\nabla v\|_{\Omega_\delta}^2 + \|v\|_{\Omega_\delta}^2 \leq e^{-\frac{\delta}{\sqrt{\varepsilon}}} (\varepsilon(\sigma \nabla v, \nabla v) + (\sigma v, v)).$$

We denote $L := \varepsilon(\sigma \nabla v, \nabla v) + (\sigma v, v)$ and obtain:

$$L = (\sigma v, v) + \varepsilon(\nabla(\sigma v), \nabla v) - \varepsilon(v \nabla \sigma, \nabla v).$$

Since $|\nabla \sigma| \leq \frac{\sigma}{\sqrt{\varepsilon}}$ we have

$$\varepsilon |(v \nabla \sigma, \nabla v)| \leq \frac{1}{2} (\varepsilon(\sigma \nabla v, \nabla v) + (\sigma v, v))$$

and

$$L \leq (\sigma v, v) + \varepsilon(\nabla(\sigma v), \nabla v) + \frac{1}{2} L.$$

Hence, absorbing terms, we obtain:

$$L \leq 2((\sigma v, v) + \varepsilon(\nabla(\sigma v), \nabla v)) = 2((\sigma v, v) - \varepsilon(\sigma v, \Delta v) + \varepsilon(\sigma v, \partial_n v)_{\partial\Omega}).$$

Using the equation (2.2.16) and the fact, that $\sigma = v = 1$ along $\partial\Omega$ we have:

$$L \leq 2\varepsilon(1, \partial_n v)_{\partial\Omega} = 2\varepsilon(1, \Delta v) = 2(1, v).$$

Due to the fact, that $\sigma(x) \geq 1$ we obtain:

$$L \leq 2(\sigma^{\frac{1}{2}}, v) \leq 2 \text{diam}(\Omega)^{\frac{1}{2}} (\sigma v, v)^{\frac{1}{2}} \leq 2 \text{diam}(\Omega)^{\frac{1}{2}} L^{\frac{1}{2}}.$$

Therefore,

$$L \leq 4 \text{diam}(\Omega).$$

This completes the proof. \square

The state equation (2.2.13) possesses an unique solution for all $q \in Q_{ad}$ and therefore there exists analogous to Example 1 a nonnegative number

$$J^* = \inf_{(u,q) \in \mathcal{A}} J(u), \quad (2.2.17)$$

where \mathcal{A} is the set of admissible elements, defined as in (2.2.6). A result of solvability of the problem (2.2.15) is given in the following proposition:

PROPOSITION 2.2.6. *Let for the optimal value J^* the following condition be fulfilled:*

$$J^* < \frac{1}{2} \min\{\|\bar{C}\|^2, \|C^\infty - \bar{C}\|^2\}, \quad (2.2.18)$$

where $C^\infty \in \mathbb{R}^{n_m}$ is given by $C_i^\infty = 1$, $i = 1 \dots n_m$. Then the problem (2.2.15) possesses a solution $(u, q) \in V \times Q_{ad}$.

PROOF. The ellipticity and the boundedness of the differential operator A of the equation (2.2.13) implies the existence of a continuously differentiable solution operator $S : Q_{ad} \rightarrow V$. Due to the convexity of the domain Ω the following regularity condition holds (see e.g. Grisvard [57]):

$$S(q) \in V \cap H^2(\Omega) \quad \forall q \in Q_{ad}. \quad (2.2.19)$$

This implies that the observation operator C is well defined on $S(Q_{ad})$, due to the Sobolev embedding theorem (see, e.g. Alt [2]). Therefore, there exists a sequence $\{(u_n, q_n)\} \subset \mathcal{A}$ with

$$\lim_{n \rightarrow \infty} J(u_n) = J^*.$$

Analog to the proof of Proposition 2.2.3 we obtain:

$$\|\nabla u_n\| \leq \frac{\|1\|^2}{4q_n}. \quad (2.2.20)$$

If the sequence $\{q_n\}$ is unbounded, then there exists a subsequence $\{q_{n_k}\} \subset Q_{ad}$ such that $q_{n_k} \rightarrow \infty$ for $k \rightarrow \infty$ and due to (2.2.12) $u_{n_k} \rightarrow 0$ in V . Hence, using the state equation (2.2.13) and the convexity of the domain Ω we obtain:

$$\|\nabla^2 u_n\| \leq \|\Delta u_n\| = \frac{1}{q_n} \|1 - u_n\| \quad (2.2.21)$$

and therefore, $u_{n_k} \rightarrow 0$ in $H^2(\Omega)$. This implies by virtue of Sobolev embedding theorem (see e.g. Alt [2]) that

$$C_i(u_{n_k}) = u_{n_k}(\xi_i) \rightarrow 0 \quad \text{for } k \rightarrow \infty, \quad i = 1 \dots n_m$$

and consequently $J(u_{n_k}) \rightarrow \frac{1}{2}\|\bar{C}\|^2$, which is a contradiction to (2.2.18). Therefore we obtain that the sequence $\{q_n\}$ is bounded and there exists $q \in \bar{Q}_{ad} = Q_{ad} \cup \{0\}$ with:

$$q_{n_k} \rightarrow q \quad \text{for } k \rightarrow \infty.$$

We consider two cases: $q = 0$ and $q \in Q_{ad}$. For the first case we set $v_k = 1 - u_{n_k}$ and $\varepsilon_k = q_{n_k}$ and obtain:

$$\begin{aligned} -\varepsilon_k \Delta v_k + v_k &= 0 && \text{in } \Omega, \\ v_k &= 1 && \text{on } \partial\Omega. \end{aligned} \quad (2.2.22)$$

Further, using Lemma 2.2.5 we have:

$$\|v_k\|_{\Omega_\delta} \leq 2 \operatorname{diam}(\Omega)^{\frac{1}{2}} e^{-\frac{\delta}{2\sqrt{\varepsilon_k}}},$$

where Ω_δ is defined as in Lemma 2.2.5. Then, using the equation (2.2.22) and the convexity of the domain Ω we obtain:

$$\|\nabla^2 v_k\|_{\Omega_\delta} \leq \|\Delta v_k\|_{\Omega_\delta} = \frac{1}{\varepsilon_k} \|v_k\|_{\Omega_\delta} \leq \frac{2}{\varepsilon_k} \operatorname{diam}(\Omega)^{\frac{1}{2}} e^{-\frac{\delta}{2\sqrt{\varepsilon_k}}}.$$

This implies that $v_k|_{\Omega_\delta} \rightarrow 0$ for $k \rightarrow \infty$ in $H^2(\Omega_\delta)$. Choosing δ small enough, we have by virtue of Sobolev theorem (see e.g. Alt [2]):

$$C_i(u_{n_k}) = u_{n_k}(\xi_i) \rightarrow 1 \quad \text{for } k \rightarrow \infty, \quad i = 1 \dots n_m$$

and consequently $J(u_{n_k}) \rightarrow \frac{1}{2} \|C^\infty - \bar{C}\|^2$, which is a contradiction with (2.2.18). Therefore, $q \in Q_{ad}$ and we complete the proof analog to Theorem 2.2.1 using the continuity of the solution operator S on $H^2(\Omega)$ due to (2.2.21). \square

2.3. Necessary and sufficient optimality conditions

In this section we reformulate the problem (2.1.1, 2.1.2) as an unconstrained optimization problem and derive the necessary and sufficient optimality conditions for it. We also introduce the Lagrange function, in order to obtain the optimality system for the constrained formulation of the problem under consideration. For the formulation of the second order optimality conditions we assume, that the operators A and C are of C^2 -class, which implies the same for the solution operator S , see Remark 2.2.2.

In the following we assume throughout, that we do not have any additional constraints on the parameters q or they are inactive in the solution of the problem (2.1.1, 2.1.2). Therefore, we set $Q_{ad} = Q$.

Using the solution operator S introduced in Proposition 2.2.1 we define the reduced observation operator $c : Q_0 \rightarrow Z$ by:

$$c(q) := C(S(q)). \tag{2.3.23}$$

This allows us to reformulate the problem under consideration as an unconstrained optimization problem with the reduced cost functional $j : Q_0 \rightarrow \mathbb{R}$:

$$\text{Minimize } j(q) := \frac{1}{2} \|c(q) - \bar{C}\|_Z^2 + \frac{\alpha}{2} \|q - \bar{q}\|_Q^2, \quad q \in Q. \tag{2.3.24}$$

The first- and second-order necessary optimality condition for this problem are formulated by the following theorem:

THEOREM 2.3.1. *Let $q \in Q$ be a local solution of the problem (2.3.24). Then there holds:*

(i) q is stationary point of j , i.e.

$$j'(q)(\delta q) = 0 \quad \forall \delta q \in Q, \quad (2.3.25)$$

(ii) $j''(q)$ is positive semidefinite, i.e.

$$j''(q)(\delta q, \delta q) \geq 0 \quad \forall \delta q \in Q. \quad (2.3.26)$$

The proof of this theorem can be found in the standard literature, see e.g. Nocedal & Wright [78].

The sufficient optimality conditions for the problem (2.3.24) are given by the following theorem:

THEOREM 2.3.2. *Let for $q \in Q$ the first order necessary condition (2.3.25) be fulfilled. Let in addition the second derivative $j''(q)$ be positive definite, i.e. there exists $\gamma \in \mathbb{R}_+$ with*

$$j''(q)(\delta q, \delta q) \geq \gamma \|\delta q\|_Q^2 \quad \forall \delta q \in Q, \quad (2.3.27)$$

then q is a local minimum of j .

The proof of this theorem can be also found in the standard literature, see e.g. Nocedal & Wright [78].

Denoting by $G = c'(q)$ the Jacobian matrix of the reduced observation operator c , the first order necessary condition (2.3.25) for the problem (2.3.24) reads:

$$G^*c(q) + \alpha q = G^*\bar{C} + \alpha\bar{q}. \quad (2.3.28)$$

In the following proposition we compute the Jacobian G .

PROPOSITION 2.3.1. *Let the reduced observation operator c be defined as in (2.3.23). Then its partial derivatives can be computed as follows:*

$$\frac{\partial c_i}{\partial q_j}(q) = G_{ij} = C'_i(u)(w_j), \quad i = 1 \dots n_m, \quad j = 1 \dots n_p, \quad (2.3.29)$$

with $u = S(q)$, C_i and c_i denote the components of the observation and the reduced observation operators respectively; G_{ij} denotes the entries of the Jacobian matrix $G = c'(q)$ and $w_j \in V$ is the solution of the following tangent problem:

$$a'_u(u, q)(w_j, \phi) = -a'_{q_j}(u, q)(1, \phi) \quad \forall \phi \in V. \quad (2.3.30)$$

PROOF. Taking the derivatives of

$$a(S(q), q)(\phi) = f(\phi) \quad \forall \phi \in V, \quad (2.3.31)$$

we obtain:

$$a'_u(u, q)(S'_{q_j}(q)(1), \phi) + a'_{q_j}(u, q)(1, \phi) = 0 \quad \forall \phi \in V, \quad (2.3.32)$$

where $S'_{q_j}(q)(1)$ denotes the partial derivative of the solution operator S with respect to q_j .

By the chain rule we have:

$$G_{ij} = \frac{\partial c_i}{\partial q_j}(q) = C'_i(u)(S'_{q_j}(q)(1)). \quad (2.3.33)$$

To complete the proof we use the definition (2.3.30) of w_j . \square

The Hessian matrix $\nabla^2 j(q)$ of the reduced cost functional j , which is as usual given by the identification

$$\langle \delta q, \nabla^2 j(q) \tau q \rangle_Q = j''(q)(\delta q, \tau q) \quad \forall \delta q, \tau q \in Q$$

can be expressed as:

$$\nabla^2 j(q) = G^* G + M + \alpha I, \quad (2.3.34)$$

where I denotes the identity matrix on Q , G^* is the transpose of G and the matrix $M \in \mathbb{R}^{n_p \times n_p}$ is defined by

$$M := - \sum_{i=1}^{n_m} c''_i(q) R_i^{LS}. \quad (2.3.35)$$

Here, R_i^{LS} denotes the i -th component of the least squares residual $R^{LS}(u)$ defined in (2.1.3) with $u = S(q)$.

We collect the necessary information for computation of M in the next proposition.

PROPOSITION 2.3.2. *The entries M_{jk} of the matrix M defined in (2.3.35) can be computed by:*

$$\begin{aligned} M_{jk} = & -a''_{uu}(u, q)(w_j, w_k, z) - a''_{uq_j}(u, q)(w_k, 1, z) - a''_{q_j q_k}(u, q)(1, 1, z) \\ & - a''_{uq_k}(u, q)(w_j, 1, z) - \langle C'''(u)(w_j, w_k), R^{LS}(u) \rangle_Z, \end{aligned}$$

where $u = S(q)$. Further, $w_j \in V$ is defined in (2.3.30) and $z \in V$ is the solution of the following adjoint equation:

$$a'_u(u, q)(\phi, z) = -\langle R^{LS}(u), C'(u)(\phi) \rangle_Z \quad \forall \phi \in V. \quad (2.3.36)$$

The proof of Proposition 2.3.2 is similar to the one of Proposition 2.3.1.

REMARK 2.3.3. The existence of the solutions of the tangent equation (2.3.30) and the adjoint equation (2.3.36) is given as a consequence of Assumption (A2).

The solution z of (2.3.36) is also called the Lagrange multiplier for the problem (2.1.1, 2.1.2). It is a main component of a so called *Lagrange principle*, which is used to establish the optimality

system for the constraint formulation of the parameter identification problem. For the problem under consideration we introduce the following Lagrange function:

$$\mathcal{L}(u, q, z) = \frac{1}{2}\|C(u) - \bar{C}\|_Z^2 + \frac{\alpha}{2}\|q - \bar{q}\|_Q^2 + f(z) - a(u, q)(z), \quad (2.3.37)$$

for $u \in V$, $q \in Q$ and $z \in V$. The derivative of \mathcal{L} is expressed by means of the residual functionals, defined, using the abbreviation $\xi := (u, q, z)$, by

$$\rho_u(\xi)(\phi) := f(\phi) - a(u, q)(\phi), \quad (2.3.38)$$

$$\rho_z(\xi)(\psi) := -\langle C'(u)(\psi), R^{LS}(u) \rangle_Z - a'_u(u, q)(\psi, z), \quad (2.3.39)$$

$$\rho_q(\xi)(\sigma) := \alpha\langle q - \bar{q}, \sigma \rangle_Q - a'_q(u, q)(\sigma, z), \quad (2.3.40)$$

where $\phi, \psi \in V$ and $\sigma \in Q$ are test functions. Now we have with $\delta\xi = (\delta u, \delta q, \delta z)$:

$$\mathcal{L}'(\xi)(\delta\xi) = \rho_u(\xi)(\delta z) + \rho_z(\xi)(\delta u) + \rho_q(\xi)(\delta q). \quad (2.3.41)$$

This leads to the following formulation of the Lagrange principle:

THEOREM 2.3.3 (Lagrange principle). *Let Assumptions (A1) and (A2) be valid and $(u, q) \in V \times Q$ be a (local) solution of the problem (2.1.1, 2.1.2). Then there exists a Lagrange multiplier $z \in V$ such that the triple (u, q, z) is a stationary point of the Lagrange function defined in (2.3.37).*

PROOF. We define the Lagrange multiplier $z \in V$ as in (2.3.36) and there remains to check that the derivative $\mathcal{L}'(u, q, z)$ vanishes. The residual functional ρ_u vanishes due to the fact, that u solves the state equation for the parameter q ; the residual functional ρ_z vanishes due to the definition of z (2.3.36) and for the term $\rho_q(\xi)(\sigma)$ we obtain analog to the proof of Proposition 2.3.1:

$$\begin{aligned} j'(q)(\sigma) &= \langle C(u) - \bar{C}, C'(u)(S'(q)(\sigma)) \rangle_Z + \alpha\langle q - \bar{q}, \sigma \rangle_Q \\ &= a'_u(u, q)(S'(q)(\sigma), z) + \alpha\langle q - \bar{q}, \sigma \rangle_Q \\ &= -a'_q(u, q)(\sigma, z) + \alpha\langle q - \bar{q}, \sigma \rangle_Q = \rho_q(\xi)(\sigma). \end{aligned}$$

The application of Theorem 2.3.1 completes the proof. \square

REMARK 2.3.4. The proof of the Lagrange principle here is simple due to Assumption (A2), which implies the existence of the Lagrange multiplier z . For the discussion of the Lagrange principle in more general situation see e.g. Tikhomirov [85].

2.4. Stability and uniqueness of the solution

In this section we introduce the concept of a *stable* solution of the parameter identification problem and analyze the main properties of it. We investigate the influence of small perturbations of the problem on the corresponding solution. Among other things we prove the local

uniqueness of stable solutions. Moreover we show, that global uniqueness cannot be expected for the solution of the problem under consideration.

We call a solution $q \in Q$ of the problem (2.3.24) stable, if the sufficient optimality condition from Theorem 2.3.2 is fulfilled, i.e. if there exists $\gamma \in \mathbb{R}_+$ with:

$$j''(q)(\delta q, \delta q) \geq \gamma \|\delta q\|_Q^2 \quad \forall \delta q \in Q. \quad (2.4.42)$$

In the following proposition we derive a sufficient condition for the stability of a solution of the problem (2.3.24).

PROPOSITION 2.4.1. *Let q be a solution of the problem (2.3.24) and G be the Jacobian matrix of the reduced observation operator c . Let moreover $\underline{\lambda}_{G^*G}, \varepsilon, \sigma \in \mathbb{R}_+$ be defined by:*

$$\underline{\lambda}_{G^*G} = \min\{\lambda \in \mathbb{R} \mid \lambda \text{ is an eigenvalue of } G^*G\},$$

$$\varepsilon = \|c(q) - \bar{C}\|_Z$$

and

$$\sigma = \sup_{\delta q \in Q} \frac{\|c''(q)(\delta q, \delta q)\|_Z}{\|\delta q\|_Q^2}.$$

Then the existence of a positive number γ with

$$\underline{\lambda}_{G^*G} - \sigma\varepsilon + \alpha \geq \gamma > 0 \quad (2.4.43)$$

is sufficient for the stability of a solution q .

PROOF. The matrix G^*G is symmetric and positive semidefinite. This implies that all eigenvalues are real and $\underline{\lambda}_{G^*G} \in \mathbb{R}_+ \cup \{0\}$.

For the second derivative of the reduced cost functional j we obtain:

$$\begin{aligned} j''(q)(\delta q, \delta q) &= \delta q^* G^* G \delta q + \langle c''(q)(\delta q, \delta q), c(q) - \bar{C} \rangle_Z + \alpha \langle \delta q, \delta q \rangle_Q \geq \\ &\geq \underline{\lambda}_{G^*G} \|\delta q\|_Q^2 - \|c''(q)(\delta q, \delta q)\|_Z \|c(q) - \bar{C}\|_Z + \alpha \|\delta q\|_Q^2 \geq \\ &\geq (\underline{\lambda}_{G^*G} - \sigma\varepsilon + \alpha) \|\delta q\|_Q^2. \end{aligned}$$

This completes the proof. □

REMARK 2.4.2. If the cost functional vanishes in the solution q , i.e. if $\varepsilon = 0$, then each of the following conditions is sufficient for the stability:

- (i) The Jacobian matrix G has full rank n_p ,
- (ii) The regularization parameter α is positive.

For a stable solution q of the problem (2.3.24) we formulate the following stability theorem, see Bock [25] for analogous argumentation:

THEOREM 2.4.1. *Let q be a stable solution of the problem (2.3.24) and $\tilde{j}(q, \epsilon)$ be a perturbed cost functional with a perturbation parameter $\epsilon \in \mathcal{E} = \mathbb{R}^k$. We assume, that $\tilde{j} : Q \times \mathcal{E} \rightarrow \mathbb{R}$ is continuously differentiable with*

$$j(q) = \tilde{j}(q, 0) \quad \forall q \in Q.$$

Then there exist a neighborhood $Q_1 \subset Q$ of q , a neighborhood $\mathcal{E}_1 \subset \mathcal{E}$ of $\epsilon = 0$ and a continuously differentiable function $f : \mathcal{E}_1 \rightarrow Q_1$ such that $f(0) = q$ and $q_\epsilon = f(\epsilon)$ is a local minimum of $\tilde{j}(q, \epsilon)$ for all $\epsilon \in \mathcal{E}_1$.

PROOF. We apply the implicit function theorem (see, e.g., Dieudonné [40]) to the function $F(q, \epsilon) = \nabla_q \tilde{j}(q, \epsilon)$, where the gradient $\nabla_q \tilde{j}(q, \epsilon) \in Q$ is as usual defined by the identification:

$$\tilde{j}'_q(q, \epsilon)(\delta q) = \langle \nabla_q \tilde{j}(q, \epsilon), \delta q \rangle_Q \quad \forall \delta q \in Q.$$

There holds:

- (i) $F(q, 0) = 0$,
- (ii) $F'_q(q, 0) = \nabla^2 j(q)$ is positive definite and consequently regular.

The application of the implicit function theorem provides the existence of the neighborhoods $\tilde{\mathcal{E}}_1 \subset \mathcal{E}$ of $\epsilon = 0$, $Q_1 \subset Q$ of q and a continuously differentiable function $f : \tilde{\mathcal{E}}_1 \rightarrow Q_1$ such that:

$$\nabla_q \tilde{j}(f(\epsilon), \epsilon) = 0 \quad \forall \epsilon \in \tilde{\mathcal{E}}_1.$$

Moreover, there exists a neighborhood $\mathcal{E}_1 \subset \tilde{\mathcal{E}}_1$ such that

$$\nabla_q^2 \tilde{j}(f(\epsilon), \epsilon) \text{ is positive definite } \forall \epsilon \in \mathcal{E}_1.$$

Therefore, we obtain by using the sufficient optimality conditions from Theorem 2.3.2, that $q_\epsilon = f(\epsilon)$ is a local minimum of $\tilde{j}(q, \epsilon)$ for all $\epsilon \in \mathcal{E}_1$. This completes the proof. \square

The above result means, that “small” (continuous) perturbations of the cost functional j lead to “small” perturbations of the solution q . In the next theorem we give a more precise result for perturbations of the measurement vector \bar{C} .

THEOREM 2.4.2. *Let q be a stable solution of the problem (2.3.24), $\delta \bar{C}$ be a perturbation of the measurement vector \bar{C} and \tilde{q} a solution of the perturbed parameter identification problem with the measurements $\bar{C} + \delta \bar{C}$. Then, there holds for $\delta q = \tilde{q} - q$:*

$$\frac{\delta q_j}{q_j} = \sum_{i=1}^{n_m} \kappa_{ij} \frac{\delta \bar{C}_i}{\bar{C}_i} + O(\|\delta \bar{C}\|_Z^2),$$

where the so called (relative) condition numbers κ_{ij} are given by:

$$\kappa_{ij} = (H^{-1}G^*)_{ji} \frac{\bar{C}_i}{q_j},$$

with the Hessian $H = \nabla^2 j(q)$ and the Jacobian matrix $G = c'(q)$.

PROOF. Similar to the proof of the Theorem 2.4.1 we define a function $F : Q \times Z \rightarrow Q$ describing the first order necessary condition for a given measurement vector \bar{C} by:

$$F(q, \bar{C}) = G^*(c(q) - \bar{C}).$$

Due to the implicit function theorem there exists a neighborhood $Z_0 \subset Z$ and a continuously differentiable function $f : Z_0 \rightarrow Q$ such that:

$$F(f(\bar{C}), \bar{C}) = 0 \quad \forall \bar{C} \in Z_0$$

and $f(\bar{C})$ is a solution for the problem (2.3.24) for the measurement vector \bar{C} . There holds:

$$F'_q(q, \bar{C}) = H$$

and

$$F'_{\bar{C}}(q, \bar{C}) = -G^*.$$

Therefore, the derivative $f'(\bar{C})$ is given by:

$$f'(\bar{C}) = H^{-1}G^*.$$

We complete the proof by using the fact, that

$$\delta q = f(\bar{C} + \delta \bar{C}) - f(\bar{C}) = f'(\bar{C})(\delta \bar{C}) + O(\|\delta \bar{C}\|_Z^2).$$

□

This theorem is a typical result of the sensitivity analysis. The relative condition number κ_{ij} describes the relative importance of the i th measurement for the parameter q_j . It can be simply computed by a post-processing of the optimization algorithm described in Chapter 4, see Becker & Vexler [23]. For the sensitivity analysis of optimal control problems, see e.g. Malanowski [75] or Griesse [55, 56].

From Theorem 2.4.2 we obtain the following corollary:

COROLLARY 2.4.3. *Let $\underline{\lambda}_H$ denote the smallest eigenvalue of the Hessian H , $\bar{\lambda}_{GG^*}$ the largest eigenvalue of the matrix GG^* and κ be defined by:*

$$\kappa = \underline{\lambda}_H^{-1} \sqrt{\bar{\lambda}_{GG^*}} \frac{\|\bar{C}\|_Z}{\|q\|_Q}.$$

Then, with the notation from Theorem 2.4.2 there holds:

$$\frac{\|\delta q\|_Q}{\|q\|_Q} \leq \kappa \frac{\|\delta \bar{C}\|_Z}{\|\bar{C}\|_Z} + O(\|\delta \bar{C}\|_Z^2).$$

PROOF. From Theorem 2.4.2 we obtain:

$$\|\delta q\| \leq \|H^{-1}G^*\delta\bar{C}\| + O(\|\delta\bar{C}\|_Z^2).$$

For the term $\|H^{-1}G^*\delta\bar{C}\|$ we have:

$$\|H^{-1}G^*\delta\bar{C}\|^2 = (G^*\delta\bar{C})^*(H^{-1})^2(G^*\delta\bar{C}) \leq \underline{\lambda}_H^{-2} \delta\bar{C}^*(GG^*)\delta\bar{C} \leq \underline{\lambda}_H^{-2} \bar{\lambda}_{GG^*} \|\delta\bar{C}\|_Z^2.$$

This completes the proof. \square

Another consequence of the stability is the local uniqueness, formulated in the following theorem:

THEOREM 2.4.3 (Local uniqueness). *Let q be a stable solution of the problem (2.3.24), then there exist a neighborhood $Q_1 \subset Q$ of q such that:*

$$j(\tilde{q}) > j(q) \quad \forall \tilde{q} \in Q_1 \setminus \{q\}.$$

PROOF. Due to the positive definiteness of $j''(q)$ there exists a neighborhood $Q_1 \subset Q$ of q such that $j''(\hat{q})$ is also positive definite for all $\hat{q} \in Q_1$. Let now $\tilde{q} \in Q_1$ be arbitrary chosen. There holds with an appropriate $\hat{q} \in Q_1$:

$$j(\tilde{q}) - j(q) = j'(q)(\tilde{q} - q) + j''(\hat{q})(\tilde{q} - q, \tilde{q} - q).$$

We complete the proof by using the first order necessary condition from Theorem 2.3.1 and the positive definiteness of $j''(\hat{q})$. \square

A natural consequence of the above theorem is that a stable solution q of the problem (2.3.24) possesses a neighborhood, which does not contain any additional (local) solution of the problem (2.3.24).

In the following example we show that the local uniqueness of a solution q of the problem (2.3.24) does not guarantee the global uniqueness. To this end, we consider a parameter identification problem, which possesses two different stable solutions.

Let $I = (0, 1)$ be the unit interval, we choose $V = H_0^1(I)$ and $Q = \mathbb{R}$. With these settings we consider the following parameter identification problem for $u \in V$ and $q \in Q$:

$$\text{Minimize } J(u) = (u(0.5) - \bar{C})^2 \tag{2.4.44}$$

under the constraint

$$\begin{aligned} -u'' + qu' &= 1 \quad \text{in } I, \\ u(0) = u(1) &= 0. \end{aligned} \tag{2.4.45}$$

The solution operator S is here given by:

$$S(q)(x) = \frac{xe^q - x - e^{qx} + 1}{q(e^q - 1)}.$$

In Figure 2.4.2 we see the solutions of the state equations for two choices of the parameter: $q_1 = 5$ and $q_2 = -5$.

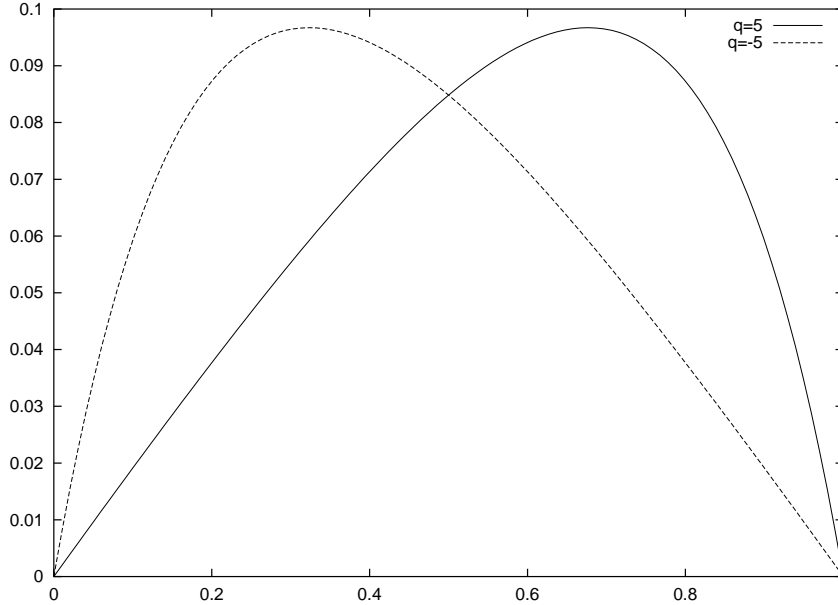


FIGURE 2.4.2. The solutions of the state equations for two choices of the parameter: $q_1 = 5$ (solid line) and $q_2 = -5$ (dashed line).

One simply obtains that

$$S(q_1)(0.5) = S(q_2)(0.5) = \frac{e^{2.5} - 1}{10(e^{2.5} + 1)}.$$

It leads to the following situation: We choose

$$\bar{C} = \frac{e^{2.5} - 1}{10(e^{2.5} + 1)}$$

and obtain, that the problem (2.4.44, 2.4.45) possesses at least two solutions $(q_1, S(q_1))$ and $(q_2, S(q_2))$. Applying Proposition 2.4.1 to this problem we obtain that these both solutions are stable with $j(q_1) = j(q_2) = 0$ (cf. Remark 2.4.2).

2.5. Statistical considerations

In this section we touch the statistical aspects of parameter identification. We give a statistical justification of using least squares method for the formulation of the parameter identification problem. Moreover, we analyze the statistical quality of the parameters obtained as a solution of the least squares problem.

By the statistical consideration of the parameter identification problem one uses the following (nonlinear) regression model (see e.g. Bates & Watts [10]):

$$\bar{C} = C(u^\tau) + \varepsilon, \quad (2.5.46)$$

where $u^\tau = S(q^\tau)$ is the solution of the state equation (2.1.1) for the *true* parameters $q^\tau \in Q$ and $\varepsilon \in Z$ denotes the measurement errors considered as a vector of random variables ε_i . We make the following standard assumption on the distribution of the measurement errors:

ASSUMPTION (S1). *The measurement errors ε_i are uncorrelated and normally distributed with the expected value $E(\varepsilon) = 0$ and the covariance matrix*

$$\text{cov}(\varepsilon) = \sigma^2 I, \quad \sigma \in \mathbb{R}_+.$$

The analysis of the general nonlinear regression model (2.5.46) requires validity of several additional assumptions on C , S , and ε , see e.g. Humak [67], and is still an active field of research. However, the special case of linear regression, i.e. if both the solution operator S and the observation operator C are linear, is well studied. In this case there holds the following theorem, see e.g. Seber & Lee [83]:

THEOREM 2.5.1. *Let Assumption (S1) be valid and $q \in Q$ be a global minimizer of the least square problem (2.3.24) (with the regularization parameter $\alpha = 0$), then among the class of linear unbiased estimates of q^τ , the estimate q is the unique one with minimum variance.*

The above theorem gives a justification of using least squares approach for parameter identification in the linear case.

The validity of Assumption (S1) allows us also to analyze the statistical quality of the solution q of the least square problem (2.3.24). This is done by introduction of so called *confidence regions*. A confidence region describes the region around the estimated parameter q , where the true parameter q^τ lies with a given probability level. Under Assumption (S1) the confidence region with the probability level $100(1 - \alpha)\%$ is given by:

$$\mathcal{R} = \{\tilde{q} \mid j(\tilde{q}) - j(q) \leq \gamma^2(\alpha)\}, \quad (2.5.47)$$

where

$$\gamma^2(\alpha) = \sigma^2 n_p F_{1-\alpha}(n_p, n_m - n_p).$$

The factor σ^2 can be estimated by:

$$b^2 = \frac{j(q)}{n_m - n_p}.$$

In general, there is no simple characterization of the confidence region (2.5.47). Therefore, we consider a linearized version of the confidence region:

$$\mathcal{R}_L = \{q + \delta q \mid j''(q)(\delta q, \delta q) \leq \gamma^2(\alpha)\}. \quad (2.5.48)$$

This is justified by the Taylor's formula giving:

$$j(\tilde{q}) - j(q) = j'(q)(\tilde{q} - q) + j''(q)(\tilde{q} - q, \tilde{q} - q) + o(\|\tilde{q} - q\|_Q^2)$$

and the fact, that

$$j'(q)(\tilde{q} - q) = 0$$

due to the necessary optimality condition from Theorem 2.3.1.

For a stable solution q of the problem (2.3.24) the linearized confidence region \mathcal{R}_L defines an ellipsoid (due to the positive of the Hessian matrix of j). In this case the linearized confidence region \mathcal{R}_L can be estimated by so called *confidence intervals*:

PROPOSITION 2.5.1. *For a stable solution q of the problem (2.3.24) the linearized confidence region \mathcal{R}_L , defined in (2.5.48), is a subset of a parallelepiped given as a product of confidence intervals for each component of q , i.e.:*

$$\mathcal{R}_L \subset \prod_{j=1}^{n_p} [q_j - \theta_j, q_j + \theta_j],$$

where

$$\theta_j = P_{ii}^{\frac{1}{2}} \gamma(\alpha)$$

with $P = (\nabla^2 j(q))^{-1}$.

PROOF. The proof is given by a direct calculation. □

The computation of the confidence intervals introduced above can be done by a simple post-processing of the optimization algorithm described in Chapter 4. This requires only the knowledge of the diagonal elements of the inverse P of the Hessian. If an approximative Hessian is used, one obtains approximative confidence intervals, see e.g. Bock [25] for the computation of confidence intervals based on the Gauß-Newton matrix.

CHAPTER 3

Finite Element Discretization

In this chapter we treat the discretization of parameter identification problems by the finite element method. In Section 3.1 we describe meshes over the computational domain and the corresponding finite element spaces. Moreover, we introduce finite element interpolation operators and discuss the estimates of the interpolation errors. In Section 3.2 the described finite element spaces are used for the discretization of the parameter identification problem (2.1.1, 2.1.2). Afterwards, in Section 3.3 we study the question of the existence of a solution of the discretized parameter identification problem. Section 3.4 is devoted to the derivation of a priori error estimates for the error in parameters due to the discretization. In the last section of this chapter we present some numerical results confirming these error estimates.

3.1. Triangulations and finite element spaces

In this section we describe the finite element meshes (triangulations) over the domain $\Omega \subset \mathbb{R}^d$ with a polygonal boundary $\partial\Omega$ and the corresponding finite element spaces. The case of non-polygonal boundary necessitates the treatment of curved cells and is neglected here.

We consider two dimensional meshes consisting of open quadrilaterals (*cells*) K which constitute a non-overlapping covering of the computational domain. The corresponding mesh is denoted by $\mathcal{T}_h = \{K\}$, where the mesh parameter h is defined as a cell-wise constant function by setting $h|_K = h_K$ and h_K is the diameter of K . Usually we use the symbol h also for the maximal cell size, i.e.

$$h = \max_{K \in \mathcal{T}_h} h_K.$$

The straight parts which make up the boundary ∂K of a cell K are called *faces*.

A mesh \mathcal{T}_h is called regular, if it fulfills the standard conditions for shape-regular finite element mesh (see e.g. Ciarlet [34] or Braess [31]), i.e. if the following properties are satisfied:

- (M1) $\bar{\Omega} = \cup_{K \in \mathcal{T}_h} \bar{K}$,
- (M2) $K_1 \cap K_2 = \emptyset$ or $K_1 = K_2$, $\forall K_1, K_2 \in \mathcal{T}_h$,
- (M3) Any face of any cell $K_1 \in \mathcal{T}_h$ is either a subset of the boundary $\partial\Omega$, or a face of another cell $K_2 \in \mathcal{T}_h$.

However, in order to ease the mesh refinement we weaken the condition (M3) and allow the cells to have nodes, which lie on midpoints of faces of neighboring cells. But at most one such

hanging node (see Figure 3.1.1) is permitted for each face.

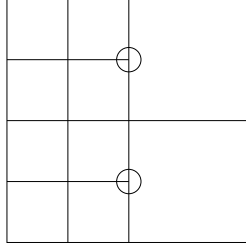


FIGURE 3.1.1. Quadrilateral mesh in two dimensions with two hanging nodes.

In addition we require that the mesh \mathcal{T}_h is organized in a patch-wise manner. This means, that it results from a coarser regular mesh \mathcal{T}_{2h} by one global refinement. By a “patch” of elements we denote a group of four cells in \mathcal{T}_h which results from a common coarser cell in \mathcal{T}_{2h} . In Figure 3.1.2, an example is shown.

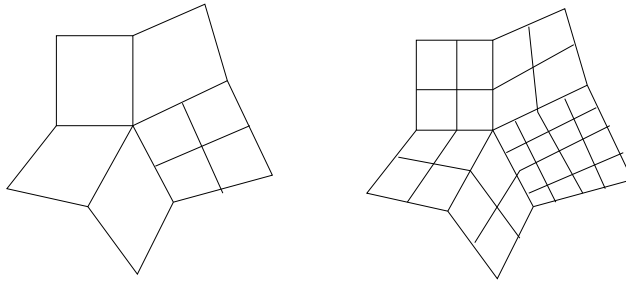


FIGURE 3.1.2. “Patch-mesh” \mathcal{T}_h (right) resulting from a coarser regular mesh \mathcal{T}_{2h} (left) by one global refinement.

On a regular mesh, we construct continuous finite element spaces $V_h \subset V$, see e.g. Ciarlet [34], Brenner & Scott [32] or Johnson [68], by:

$$V_h = \{v \in V \mid v|_K \in P(K), K \in \mathcal{T}_h\},$$

where $P(K)$ denotes a suitable space of polynomial-like functions defined on the cell $K \in \mathcal{T}_h$. In our numerical results discussed below, we mostly use (isoparametric) “bilinear” or “biquadratic” finite elements on quadrilateral meshes. In this case $P(K) = \tilde{Q}_r(K)$ ($r = 1, 2$) consists of shape functions obtained via a bilinear transformation from the space of bilinear or biquadratic functions $Q_r(\hat{K})$ on the reference cell $\hat{K} = (0, 1)^2$, where

$$Q_1(\hat{K}) = \text{span}\{1, x_1, x_2, x_1x_2\} \quad \text{or} \quad Q_2(\hat{K}) = Q_1(\hat{K}) \oplus \text{span}\{x_1^2, x_2^2, x_1x_2^2, x_2x_1^2, x_1^2x_2^2\}.$$

The case of hanging nodes requires some additional remarks. There are no degrees of freedom corresponding to these irregular nodes and the value of the finite element function is

determined by point-wise interpolation. This implies continuity and therefore global conformity. For implementation details see e.g. Carey & Oden [33].

REMARK 3.1.1. The space V_{2h} corresponding to the mesh \mathcal{T}_{2h} is defined in a similar way. The patch-structure of the mesh will be used for defining stabilization terms, see Chapter 6, and for evaluation of a posteriori error estimators, see Chapter 5.

The mesh \mathcal{T}_h is called *quasi-uniform*, if the following additional condition is valid for a moderate constant C_m :

$$(M4) \quad \frac{h_K}{\rho_K} \leq C_m \quad \forall K \in \mathcal{T}_h,$$

where ρ_K denotes the diameter of the biggest ball inscribed in K .

The approximation properties of finite element spaces can be characterized by error estimates for interpolation errors. Throughout we use two types of interpolation operators $i_h : V \rightarrow V_h$: point-wise interpolation for continuous functions and a generalized interpolation for functions in $H^1(\Omega)$, see Clément [35].

In the following proposition we collect some estimates of interpolation errors.

PROPOSITION 3.1.2. *Let \mathcal{T}_h be a quasi-uniform mesh and V_h a space of (isoparametric) bilinear finite elements, then there exists a constant C_I depending on C_m such that there holds for $u \in H^2(\Omega)$:*

$$\|u - i_h u\|_{\Omega} \leq C_I h^2 \|\nabla^2 u\|_{\Omega},$$

and

$$\|\nabla(u - i_h u)\|_{\Omega} \leq C_I h \|\nabla^2 u\|_{\Omega}.$$

The proof is given e.g. in Braess [31].

REMARK 3.1.3. For the analogous results in the case of biquadratic finite elements, see e.g. Braess [31].

3.2. Discretization of a parameter identification problem

In this section we describe the discretization of the problem (2.1.1, 2.1.2) by finite element methods. To this end, the state equation (2.1.1) is discretized using finite element spaces introduced in Section 3.1. Thereafter, we turn the discretized problem in an unconstrained formulation and derive the corresponding necessary and sufficient optimality condition.

For a given finite element space V_h the corresponding discrete solution $(u_h, q_h) \in V_h \times Q$ is determined by:

$$\text{Minimize } J(u_h, q_h) \tag{3.2.49}$$

under the constraint

$$a(u_h, q_h)(\phi_h) = f(\phi_h) \quad \forall \phi_h \in V_h, \tag{3.2.50}$$

where J is the cost functional defined in (2.1.2).

REMARK 3.2.1. Due to the finite dimension of Q , we suppose the parameter q_h in (3.2.49) to be searched in the same space Q .

We assume the discrete analog of Assumptions (A1) and (A2), in order to guarantee the existence of a continuously differentiable discrete solution operator $S_h : Q_0 \rightarrow V_h$ such that there holds for all $q_h \in Q_0$:

$$a(S_h(q_h), q_h)(\phi_h) = f(\phi_h) \quad \forall \phi \in V_h.$$

Analogous to Section 2.3, we introduce the discrete reduced observation operator $c_h : Q_0 \rightarrow Z$ by

$$c_h(q_h) = C(S_h(q_h)), \quad (3.2.51)$$

and turn the discrete problem (3.2.49, 3.2.50) into an unconstrained minimization problem:

$$\text{Minimize } j_h(q_h) := \frac{1}{2} \|c_h(q_h) - \bar{C}\|_Z^2 + \frac{\alpha}{2} \|q_h - \bar{q}\|_Q^2, \quad q_h \in Q. \quad (3.2.52)$$

Analogous to Theorem 2.3.1 the first necessary condition for this problem reads:

$$j'_h(q_h) = 0.$$

Denoting $G_h = c'_h(q_h)$ this condition can be rewritten as:

$$G_h^* c_h(q_h) + \alpha q_h = G_h^* \bar{C} + \alpha \bar{q}. \quad (3.2.53)$$

The matrix G_h can be computed analog to the Jacobian G of the reduced observation operator c , i.e there holds the following proposition:

PROPOSITION 3.2.2. *Let the discrete reduced observation operator c_h be defined by (3.2.51). Then its partial derivatives can be computed as follows:*

$$\frac{\partial c_{i,h}}{\partial q_j}(q_h) = G_{ij,h} = C'_i(u_h)(w_{j,h}), \quad i = 1 \dots n_m, \quad j = 1 \dots n_p, \quad (3.2.54)$$

with $u_h = S_h(q_h)$ and $c_{i,h}$ denote the components of the discrete reduced observation operator; $G_{ij,h}$ denotes the entries of the Jacobian matrix $G_h = c'_h(q_h)$ and $w_{j,h} \in V_h$ is the solution of the following tangent problem:

$$a'_u(u_h, q_h)(w_{j,h}, \phi_h) = -a'_{q_j}(u_h, q_h)(1, \phi_h) \quad \forall \phi_h \in V_h. \quad (3.2.55)$$

The proof of Proposition 3.2.2 is similar to the one of Proposition 2.3.1.

The sufficient optimality conditions for the discretized problem (3.2.52) are given by the application of Theorem 2.3.2 to the discrete reduced cost functional j_h :

THEOREM 3.2.1. *Let for $q_h \in Q$ the first order necessary condition (3.2.53) be fulfilled. Let in addition the second derivative $j_h''(q_h)$ be positive definite, i.e. there exists $\gamma \in \mathbb{R}_+$ with*

$$j_h''(q_h)(\delta q_h, \delta q_h) \geq \gamma \|\delta q_h\|_Q^2 \quad \forall \delta q_h \in Q, \quad (3.2.56)$$

then q_h is a local minimum of j_h .

The second derivatives of j_h , which are involved in (3.2.56), can be computed similar to the derivatives of j in Proposition 2.3.2.

PROPOSITION 3.2.3. *The Hessian matrix $\nabla^2 j_h(q_h)$ of the discrete reduced cost functional j_h is given by:*

$$\nabla^2 j_h(q_h) = G_h^* G_h + M_h + \alpha I, \quad (3.2.57)$$

where the entries $M_{jk,h}$ of the matrix M_h can be computed by:

$$\begin{aligned} M_{jk,h} = & -a''_{uu}(u_h, q_h)(w_{j,h}, w_{k,h}, z_h) - a''_{uq_j}(u_h, q_h)(w_{k,h}, 1, z_h) - a''_{q_j q_k}(u_h, q_h)(1, 1, z_h) \\ & - a''_{uq_k}(u_h, q_h)(w_{j,h}, 1, z_h) - \langle C''(u_h)(w_{j,h}, w_{k,h}), R^{LS}(u_h) \rangle_Z, \end{aligned}$$

with $u_h = S_h(q_h)$. Further, $w_{j,h} \in V$ is defined in (3.2.55) and $z_h \in V_h$ is the solution of the following adjoint equation:

$$a'_u(u_h, q_h)(\phi_h, z_h) = -\langle R^{LS}(u_h), C'(u_h)(\phi_h) \rangle_Z \quad \forall \phi_h \in V_h. \quad (3.2.58)$$

The proof of this Proposition is given similar to the one of Proposition 2.3.1.

3.3. Existence of a solution of the discrete problem

In this section we analyze the question of the existence of a solution for the discrete problem (3.2.52). We prove, that under some natural assumptions on the reduced cost functional j and the discrete reduced cost functional j_h , the existence of a stable solution q of the continuous problem (2.3.24) implies the existence of a stable solution q_h of the problem (3.2.52).

Considering the discretization of the problem (2.3.24) as a special perturbation of it, the question of the existence of a solution resembles the same question for a perturbed problem, see Section 2.4. However, the dependence of the discrete reduced cost functional j_h on the discretization parameter h is not described by a continuous differentiable function such that we cannot apply Theorem 2.4.1 in this case. For this reason we use here another approach based on the following theorem.

THEOREM 3.3.1. *Let F and F_h ($h \in \mathbb{R}_+$) be continuously differentiable operators*

$$F, F_h : \mathbb{R}^n \rightarrow \mathbb{R}^n$$

and $x \in \mathbb{R}^n$ be a solution of $F(x) = 0$. Moreover let the following conditions be fulfilled:

(i) The derivative $F'(x)$ is positive definite, i.e. there exists a constant $\gamma > 0$ with

$$p^* F'(x) p \geq \gamma \|p\|^2 \quad \forall p \in \mathbb{R}^n,$$

(ii)

$$\lim_{h \rightarrow 0} \|F(x) - F_h(x)\| = 0,$$

(iii)

$$\lim_{h \rightarrow 0} \|F'(x) - F'_h(x)\| = 0,$$

(iv) There is a neighborhood U of x and a positive number $L \in \mathbb{R}_+$ such that

$$\|F'_h(\xi) - F'_h(\eta)\| \leq L \|\xi - \eta\| \quad \forall \xi, \eta \in U \quad \forall h \in \mathbb{R}_+.$$

Then, for h small enough there exists $x_h \in U$ such that $F_h(x_h) = 0$ and $F'_h(x_h)$ is positive definite uniformly in h .

PROOF. Due to the condition (iii) we can choose a positive number $h_1 \in \mathbb{R}_+$ such that there holds for $h \leq h_1$:

$$\|F'(x) - F'_h(x)\| \leq \frac{\gamma}{4}. \quad (3.3.59)$$

Moreover we choose a positive number $\delta \leq \frac{\gamma}{4L}$ such that

$$B_\delta(x) = \{\xi \mid \|x - \xi\| \leq \delta\} \subset U. \quad (3.3.60)$$

For this choice we obtain, that $F'_h(\xi)$ is positive definite in $B_\delta(x)$ for $h \leq h_1$ uniformly in h :

$$\begin{aligned} p^* F'_h(\xi) p &= p^* F'(x) p + p^* (F'_h(x) - F'(x)) p + p^* (F'_h(\xi) - F'_h(x)) p \\ &\geq \gamma \|p\|^2 - \|F'_h(x) - F'(x)\| \|p\|^2 - \|F'_h(\xi) - F'_h(x)\| \|p\|^2 \\ &\geq (\gamma - \frac{\gamma}{4} - L\delta) \|p\|^2 \geq \frac{\gamma}{4} \|p\|^2. \end{aligned} \quad (3.3.61)$$

Moreover we obtain for $\xi \in B_\delta(x)$ and $h \leq h_1$ that the derivative $F'_h(\xi)$ is bounded uniformly in h , i.e

$$\|F'_h(\xi)\| \leq \beta := \|F'(x)\| + \frac{\gamma}{2}. \quad (3.3.62)$$

Next, we prove, that there exists a unique $x_h \in B_\delta(x)$ with $F_h(x_h) = 0$. To this end, we define an operator $D_s : \mathbb{R}^n \rightarrow \mathbb{R}^n$ by

$$D_s(\xi) = \xi - s F_h(\xi)$$

with $s \in \mathbb{R}_+$. For a certain choice of s , we show, that D_s is a contraction on $B_\delta(x)$ and use the Banach fixed point theorem.

There holds for $\xi \in B_\delta(x)$, $h \leq h_1$ and an arbitrary $p \in \mathbb{R}^n$:

$$\begin{aligned} \|D'_s(\xi) p\|^2 &= \|p - s F'_h(\xi) p\|^2 = \|p\|^2 - 2s p^* F'_h(\xi) p + s^2 \|F'_h(\xi) p\|^2 \\ &\leq (1 - s \frac{\gamma}{2} + s^2 \beta^2) \|p\|^2. \end{aligned}$$

For the choice $s = \frac{\gamma}{4\beta^2}$ there holds:

$$\|D'_s(\xi)p\|^2 \leq \left(1 - \frac{\gamma^2}{16\beta^2}\right)\|p\|^2$$

and consequently

$$\|D'_s(\xi)\| \leq \left(1 - \frac{\gamma^2}{16\beta^2}\right)^{1/2} < 1.$$

Moreover, for an arbitrary $\xi \in B_r(x)$ there holds:

$$\begin{aligned} \|x - D_s(\xi)\| &= \|D_s(x) - D_s(\xi) + sF_h(x)\| \\ &\leq \|D_s(x) - D_s(\xi)\| + s\|F_h(x) - F(x)\| \leq \|D'_s(\eta)\|\|x - \xi\| + s\|F_h(x) - F(x)\|, \end{aligned}$$

for a certain $\eta \in B_r$. Therefore, there holds:

$$\|x - D_s(\xi)\| \leq \left(1 - \frac{\gamma^2}{16\beta^2}\right)^{1/2}\delta + s\|F_h(x) - F(x)\|.$$

Due to the condition (ii) there is a number $h_2 \in \mathbb{R}_+$ such that there holds for $h \leq h_2$:

$$\|F_h(x) - F(x)\| \leq \frac{(1 - (1 - \frac{\gamma^2}{16\beta^2})^{1/2})\delta}{s}.$$

This implies

$$\|x - D_s(\xi)\| \leq \delta.$$

for $h \leq h_2$ and consequently $D_s(\xi) \in B_r(x)$.

We define h_0 by

$$h_0 := \min(h_1, h_2)$$

and obtain by the Banach fixed point theorem the existence of $x_h \in B_\delta(x)$ with $F_h(x_h) = 0$ for $h \leq h_0$. Due to the construction (3.3.60) of $B_\delta(x)$ the derivative $F'_h(x_h)$ is positive definite with a constant $\frac{\gamma}{4}$, see (3.3.61), i.e. uniformly in h . This completes the proof. \square

REMARK 3.3.1. The Lipschitz condition (iv) can be weakened in the following sense: The constant L may depend on the discretization parameter h , i.e. $L = L(h)$. However one should require, that:

$$\lim_{h \rightarrow 0} L(h)\|F(x) - F_h(x)\| = 0.$$

Then the existence of a discrete solution x_h is ensured as well. This can be proved similarly to Theorem 3.3.1.

From the above theorem we derive a sufficient condition for the existence of a solution of the problem (3.2.52) in the following corollary:

COROLLARY 3.3.2. *Let $q \in Q$ be a stable solution of (2.3.24). Let moreover, the following conditions for the reduced cost functional j and the discrete reduced cost functional j_h be fulfilled:*

(i)

$$\lim_{h \rightarrow 0} \|j'(q) - j'_h(q)\| = 0,$$

(ii)

$$\lim_{h \rightarrow 0} \|j''(q) - j''_h(q)\| = 0.$$

(iii) *There is a neighborhood U of q and a positive number $L \in \mathbb{R}_+$ such that*

$$\|j''_h(\xi) - j''_h(\eta)\| \leq L\|\xi - \eta\| \quad \forall \xi, \eta \in U \quad \forall h \in \mathbb{R}_+.$$

Then, for h small enough there exists a solution $q_h \in Q$ of the discrete problem (3.2.52), which satisfies the sufficient optimality condition from Theorem 3.2.1. Moreover, the second derivative $j''_h(q_h)$ is positive definite uniformly in h .

PROOF. The proof is given by the application of Theorem 3.3.1 on the functions

$$F(q) = \nabla j(q), \quad F_h(q) = \nabla j_h(q).$$

□

The approximation conditions (i) and (ii) and the Lipschitz condition (iii) from Corollary 3.3.2 are valid for a big class of parameter identification problems. Moreover, this result can be used not only in the context of finite element discretization. In the next section we check the conditions from Corollary 3.3.2 for a class of parameter identification problem and use it for derivation of the corresponding a priori error estimates.

3.4. A priori error analysis

The aim of this section is the derivation of the a priori error estimates for the error $q - q_h$ between the solution of the continuous parameter identification problem (2.3.24) and the discretized problem (3.2.52) obtained by the finite element discretization, see Section 3.2. There is a number of publications, where a priori error estimates are derived for optimal control problem governed by partial differential equations, see e.g. Falk [48], Arada, Casas & Tröltzsch [3], Deckelnic & Hinze [37] and Gunzburger & Hou [60]. However, to our knowledge, there are only few published results on this topic in the context of parameter identification problems, see Falk [49], Neittaanmaeki & Tai [77] and Kärkkäinen [69]. For the discussion of this subject in the case of ODE constrained parameter identification see Bock [25].

In the following theorem, we derive a general a priori error estimate for the problem under consideration, based on Theorem 3.3.1 and Corollary 3.3.2.

THEOREM 3.4.1. *Let $q \in Q$ be a stable solution of (2.3.24) and the conditions from Corollary 3.3.2 be valid. Then, for h small enough, there holds the following a priori error estimate:*

$$\|q - q_h\| \leq C_s \|j'(q) - j'_h(q)\|,$$

where the constant C_s does not depend on the discretization parameter h .

PROOF. By virtue of Corollary 3.3.2 the existence of q_h is ensured. Moreover, there exists a ball $B_\delta(q)$ (cf. (3.3.60) in the proof of Theorem 3.3.1) such that $q_h \in B_\delta(q)$ and the second derivative $j_h''(\xi)$ is positive definite for all $\xi \in B_\delta(q)$ uniformly in h . Therefore, there exists a constant $\gamma \in \mathbb{R}_+$ and $h_0 \in \mathbb{R}_+$ such that there holds:

$$j_h''(\xi)(p, p) \geq \gamma \|p\|_Q^2 \quad \forall \xi \in B_\delta(q) \quad \forall h \leq h_0.$$

Using this expression we obtain:

$$j_h'(q)(q - q_h) - j_h'(q_h)(q - q_h) = j_h''(\xi)(q - q_h, q - q_h) \geq \gamma \|q - q_h\|_Q^2,$$

for a certain $\xi \in B_\delta(q)$. Hence, using the necessary optimality conditions for the continuous and the discrete problem, we obtain:

$$\begin{aligned} \|q - q_h\|_Q^2 &\leq \frac{1}{\gamma} |j_h'(q)(q - q_h) - j_h'(q_h)(q - q_h)| = \frac{1}{\gamma} |j_h'(q)(q - q_h) - j'(q)(q - q_h)| \\ &\leq \frac{1}{\gamma} \|j'(q) - j_h'(q)\| \|q - q_h\|_Q. \end{aligned}$$

We complete the proof by setting $C_s = \frac{1}{\gamma}$. □

This result turns out, that the asymptotical behavior of the error in parameter is driven by the consistence error $\|j'(q) - j_h'(q)\|$. In the following we estimate this consistence error for parameter identification problem governed by elliptic partial differential equations of second order. For simplicity, we concentrate on the following problem:

Let $\Omega \subset \mathbb{R}^2$ be a polygonal domain, $V = H_0^1(\Omega)$ and the parameter identification problem be formulated as follows:

$$\text{Minimize } J(u) = \frac{1}{2} \|C(u) - \bar{C}\|_Z^2, \quad (3.4.63)$$

under the constraint:

$$(A(q)\nabla u, \nabla \phi) = (f, \phi) \quad \forall \phi \in V, \quad (3.4.64)$$

where $C : V \rightarrow Z = \mathbb{R}^{n_m}$ is a linear operator, $\bar{C} \in Z$ denotes a vector of measurements and $f \in L^2(\Omega)$ is a given right-hand side. The entries $a_{ij} : Q \rightarrow L^\infty(\Omega)$ of the symmetric 2×2 matrix $A(q) = (a_{ij}(q))$ are assumed to be three times continuously differentiable. For this problem we assume the existence of a stable solution $q \in Q = \mathbb{R}^{n_p}$ such that $A(q)$ is positive definite. We discretize the problem (3.4.63, 3.4.64) by bilinear finite elements on the space V_h , see Section 3.1, and prove, that the error in parameters has the same asymptotical behavior as the discretization error of the state equation with respect to the observation operator for fixed parameter q , i.e. as the error $C(S(q) - S_h(q))$. For simplicity, we first consider a regular case, where both, the state variable and the observation operator, possess a certain regularity described by the following assumptions:

ASSUMPTION (R1). Let q be a solution of (3.4.63, 3.4.64) and $g \in L^2(\Omega)$. Then for the solution $v \in V$ of the equation

$$(A(q)\nabla v, \nabla \phi) = (g, \phi) \quad \forall \phi \in V,$$

there holds: $v \in H^2(\Omega) \cap V$.

REMARK 3.4.1. Sufficient conditions for this assumption on the interior angles of $\partial\Omega$ can be found, e.g. in Grisvard [57].

In addition, we make an assumption on the regularity of the observation operator C :

ASSUMPTION (R2). Let the dual solutions $z_i \in V$, $i = 1, 2, \dots, n_m$, be defined by:

$$(A(q)\nabla \phi, \nabla z_i) = C_i(\phi) \quad \forall \phi \in V. \quad (3.4.65)$$

Then, there holds: $z_i \in H^2(\Omega) \cap V$.

The next theorem provides an a priori error estimate for the problem (3.4.63, 3.4.64).

THEOREM 3.4.2. Let for the problem (3.4.63, 3.4.64) assumptions (R1) and (R2) be satisfied, then there exists $h_0 \in \mathbb{R}_+$ such that for all $h \leq h_0$ there exists a discrete solution (u_h, q_h) and the following a priori error estimate holds:

$$\|q - q_h\|_Q = O(h^2).$$

For the proof of this theorem, we make the following preparations:

The positive definiteness of the matrix $A(q)$ implies the existence of a neighborhood Q_1 of the solution q and a positive number $\gamma \in \mathbb{R}_+$ such that there holds:

$$p^* A(\xi) p \geq \gamma \|p\|^2 \quad \forall p \in \mathbb{R}^d, \quad \forall \xi \in Q_1.$$

Due to the fact, that $A(q)$ is three times continuously differentiable, there exist the numbers $m_1, m_2, m_3 \in \mathbb{R}_+$ defined by:

$$\begin{aligned} m_1 &:= \max_j \sup_{\xi \in Q_1} \|A'_{q_j}(\xi)\|, \\ m_2 &:= \max_{j,k} \sup_{\xi \in Q_1} \|A''_{q_j q_k}(\xi)\|, \\ m_3 &:= \max_{j,k,l} \sup_{\xi \in Q_1} \|A'''_{q_j q_k q_l}(\xi)\|. \end{aligned} \quad (3.4.66)$$

We define $\bar{u}_h \in V_h$ by $\bar{u}_h = S_h(q)$, i.e. as a solution of

$$(A(q)\nabla \bar{u}_h, \nabla \phi_h) = (f, \phi_h) \quad \forall \phi_h \in V_h, \quad (3.4.67)$$

then due to the Assumption (R1) and Assumption (R2) we obtain the following a priori error estimates:

LEMMA 3.4.2. *Let u be a solution of (3.4.64), \bar{u}_h be defined as in (3.4.67) and $z_i \in V$ as in (3.4.65), then there holds:*

- (a) $\|\nabla(u - \bar{u}_h)\| \leq ch \|\nabla^2 u\|,$
- (b) $\|u - \bar{u}_h\| \leq ch^2 \|\nabla^2 u\|,$
- (c) $C_i(u - \bar{u}_h) \leq ch^2 \|\nabla^2 u\| \|\nabla^2 z_i\|.$

PROOF. For proof of (a) and (b) one uses standard techniques, see e.g. Braess [31] or Großmann & Roos [58]. For (c) we obtain:

$$\begin{aligned} C_i(u - \bar{u}_h) &= (A(q)\nabla(u - \bar{u}_h), \nabla z_i) = (A(q)\nabla(u - \bar{u}_h), \nabla(z_i - i_h z_i)) \\ &\leq m_1 \|\nabla(u - \bar{u}_h)\| \|\nabla(z_i - i_h z_i)\|. \end{aligned}$$

We complete the proof using (a), an estimate of the interpolation error, see Proposition 3.1.2, and the fact, that $z_i \in H^2(\Omega)$ due to Assumption (R2). \square

Next, we prove some a priori error estimates for the tangent solutions w_j defined in (2.3.30):

LEMMA 3.4.3. *Let $w_j \in V$ be defined as in (2.3.30) and $\bar{w}_{j,h} \in V_h$ be defined by:*

$$(A(q)\nabla\bar{w}_{j,h}, \nabla\phi_h) = -(A'_{q_j}(q)\nabla\bar{u}_h, \nabla\phi_h) \quad \forall\phi_h \in V_h. \quad (3.4.68)$$

Then, there holds:

- (a) $\|\nabla(w_j - \bar{w}_{j,h})\| = O(h),$
- (b) $C_i(w_j - \bar{w}_{j,h}) = O(h^2).$

PROOF. We introduce $\tilde{w}_{j,h} \in V_h$ as a solution of:

$$(A(q)\nabla\tilde{w}_{j,h}, \nabla\phi_h) = -(A'_{q_j}(q)\nabla u, \nabla\phi_h) \quad \forall\phi_h \in V_h.$$

With the help of them, we split the error $e_w = w_j - \bar{w}_{j,h}$ in two parts: $e_w = e_1 + e_2$, where

$$e_1 = w_j - \tilde{w}_{j,h}$$

and

$$e_2 = \tilde{w}_{j,h} - \bar{w}_{j,h}.$$

First, we obtain by the same arguments as in Lemma 3.4.2, that:

$$\|\nabla e_1\| = O(h)$$

and

$$C_i(e_1) = O(h^2).$$

Next, we show the same for e_2 . For $e_2 \in V_h$ there holds a perturbed Galerkin orthogonality relation:

$$(A(q)\nabla e_2, \nabla \phi_h) = -(A'_{q_j}(q)\nabla(u - \bar{u}_h), \nabla \phi_h) \quad \forall \phi_h \in V_h. \quad (3.4.69)$$

By setting $\phi_h = e_2$ we obtain:

$$\gamma \|\nabla e_2\|^2 \leq m_1 \|\nabla(u - \bar{u}_h)\| \|\nabla e_2\|$$

and therefore:

$$\|\nabla e_2\| \leq \frac{m_1}{\gamma} \|\nabla(u - \bar{u}_h)\| = O(h).$$

For the error with respect to C_i we obtain:

$$C_i(e_2) = (A(q)\nabla e_2, \nabla z_i) = (A(q)\nabla e_2, \nabla(z_i - i_h z_i)) + (A(q)\nabla e_2, \nabla i_h z_i).$$

Due to the perturbed Galerkin orthogonality relation (3.4.69) we have:

$$\begin{aligned} C_i(e_2) &= (A(q)\nabla e_2, \nabla(z_i - i_h z_i)) - (A'_{q_j}(q)\nabla(u - \bar{u}_h), \nabla i_h z_i) \\ &= (A(q)\nabla e_2, \nabla(z_i - i_h z_i)) - (A'_{q_j}(q)\nabla(u - \bar{u}_h), \nabla z_i) + (A'_{q_j}(q)\nabla(u - \bar{u}_h), \nabla(z_i - i_h z_i)). \end{aligned}$$

For the first and third term we simply obtain asymptotical behavior $O(h^2)$ and for the term $(A'_{q_j}(q)\nabla(u - \bar{u}_h), \nabla z_i)$ we obtain by integration by parts:

$$\begin{aligned} |(A'_{q_j}(q)\nabla(u - \bar{u}_h), \nabla z_i)| &= |(u - \bar{u}_h, \nabla \cdot (A'_{q_j}(q)\nabla z_i))| \\ &\leq m_1 \|u - \bar{u}_h\| \|\nabla^2 z_i\| = O(h^2). \end{aligned}$$

Here, we used the fact, that $z_i \in H^2(\Omega)$ due to the Assumption (R2). This completes the proof. \square

The next lemma provides an a priori error estimate for the dual solution z defined in (2.3.36).

LEMMA 3.4.4. *Let the dual solution $z \in V$ be defined as in (2.3.36) and $\bar{z}_h \in V_h$ be given by:*

$$(A(q)\nabla \phi_h, \nabla \bar{z}_h) = -\langle R^{LS}(\bar{u}_h), C(\phi_h) \rangle_Z. \quad (3.4.70)$$

Then there holds:

$$\|\nabla(z - \bar{z}_h)\|_{L^2(\Omega)} = O(h).$$

PROOF. The proof of this lemma is similar to the one of Lemma 3.4.3. \square

The next lemma provides the validity of the Lipschitz condition (iii) from Corollary 3.3.2 for the problem (3.4.63, 3.4.64). For two arbitrary parameter sets $\xi, \eta \in Q_1$ we denote $u_\xi = S_h(\xi)$ and $u_\eta = S_h(\eta)$. Correspondingly, we define $w_{j,\xi} \in V_h$ and $z_\xi \in V_h$ similar to $\bar{w}_{j,h}$ and \bar{z}_h for $q = \xi$ by:

$$(A(\xi)\nabla w_{j,\xi}, \nabla \phi_h) = -(A'_{q_j}(\xi)\nabla u_\xi, \nabla \phi_h) \quad \forall \phi_h \in V_h$$

and

$$(A(\xi)\nabla\phi_h, \nabla z_\xi) = -\langle R^{LS}(u_\xi), C(\phi_h) \rangle_Z.$$

Moreover, we define $w_{j,\eta} \in V_h$ and $z_\eta \in V_h$ in the similar way.

LEMMA 3.4.5. *For the above definition of u_ξ , u_η , $w_{j,\xi}$, $w_{j,\eta}$, z_ξ and z_η we obtain:*

$$(a) \quad \|\nabla(u_\xi - u_\eta)\| \leq L_1 \|\xi - \eta\|_Q$$

$$(b) \quad \|\nabla(w_{j,\xi} - w_{j,\eta})\| \leq L_2 \|\xi - \eta\|_Q$$

$$(c) \quad \|\nabla(z_\xi - z_\eta)\| \leq L_3 \|\xi - \eta\|_Q,$$

where L_1 , L_2 and L_3 depend only on the domain Ω , the observation operator C , constants m_1 , m_2 , γ , and additionally on the data \bar{C} and f .

PROOF. Due to the definition of u_ξ and u_η we obtain:

$$(A(\xi)\nabla(u_\xi - u_\eta), \nabla\phi_h) + ((A(\xi) - A(\eta))\nabla u_\eta, \nabla\phi_h) = 0 \quad \forall \phi_h \in V_h.$$

By setting $\phi_h = u_\xi - u_\eta$ we obtain:

$$\begin{aligned} \gamma \|\nabla(u_\xi - u_\eta)\|^2 &\leq (A(\xi)\nabla(u_\xi - u_\eta), \nabla(u_\xi - u_\eta)) \\ &= -((A(\xi) - A(\eta))\nabla u_\eta, \nabla(u_\xi - u_\eta)) \\ &\leq m_1 \|\xi - \eta\|_Q \|\nabla u_\eta\| \|\nabla(u_\xi - u_\eta)\|. \end{aligned} \tag{3.4.71}$$

It remains to show, that $\|\nabla u_\eta\|$ is bounded uniformly in h and in $\eta \in Q_1$. To this end, we use the positive definiteness of the matrix $A(\eta)$ and obtain:

$$\gamma \|\nabla u_\eta\|^2 \leq (A(\eta)\nabla u_\eta, \nabla u_\eta) = (f, u_\eta) \leq \|f\| \|u_\eta\|.$$

Due to the Poincare inequality with the Poincare constant C_p we obtain

$$\|\nabla u_\eta\| \leq \frac{C_p}{\gamma} \|f\|.$$

Therefore, we use (3.4.71) and conclude, that

$$\|\nabla(u_\xi - u_\eta)\| \leq m_1 \frac{C_p}{\gamma} \|f\| \|\xi - \eta\|_Q.$$

The estimates (b) and (c) are obtained in a similar way. \square

Next, we prove Theorem 3.4.2.

PROOF. First, we check the conditions (i), (ii) and (iii) from Corollary 3.3.2. There holds:

$$\begin{aligned} \frac{\partial}{\partial q_j} j(q) - \frac{\partial}{\partial q_j} j_h(q) &= \langle C(u) - \bar{C}, C(w_j) \rangle_Z - \langle C(\bar{u}_h) - \bar{C}, C(\bar{w}_{j,h}) \rangle_Z \\ &= \langle C(u) - \bar{C}, C(w_j - \bar{w}_{j,h}) \rangle_Z + \langle C(u - \bar{u}_h), C(\bar{w}_{j,h}) \rangle_Z. \end{aligned}$$

Using Lemma 3.4.2 and Lemma 3.4.3 and the fact, that $\|\nabla \bar{w}_{j,h}\|$ is bounded uniformly in h we obtain:

$$\|j'(q) - j'_h(q)\| = O(h^2).$$

Analogous we obtain the condition (ii) from Corollary 3.3.2 using Lemma 3.4.2, Lemma 3.4.3, Lemma 3.4.4 and the representation of the Hessians $\nabla^2 j(q)$ from Proposition 2.3.2 and $\nabla^2 j_h(q)$ from Proposition 3.2.3.

It remains to check the condition (iii) from Corollary 3.3.2. It is obtained using Lemma 3.4.5, the fact, that $\|\nabla \bar{w}_{j,h}\|$ and $\|\nabla \bar{z}_h\|$ are bounded uniformly in h and the representation of the Hessian of $\nabla^2 j_h(q)$ from Proposition 3.2.3. The application of Theorem 3.4.1 completes the proof. \square

Theorem 3.4.2 is proved on quite restrictive assumptions (R1) and (R2), but it can be simply extended for the case:

$$u \in W^{2,s}(\Omega) \quad z_i \in W^{2,t}(\Omega)$$

where

$$1 < s, t < \infty, \quad \frac{1}{s} + \frac{1}{t} = 1.$$

However, this does not include the case of point-wise observation. For this case, the following result holds:

THEOREM 3.4.3. *Let $q \in Q$ be a stable solution of (3.4.63, 3.4.64) with point-wise observations, i.e. $C_i(v) = v(\xi_i)$ with given points ξ_i in a convex domain Ω . Then, for h small enough, there exist a stable solution $q_h \in Q$ of the discretized problem, and there holds the following a priori error estimate:*

$$\|q - q_h\|_Q \leq C_s |\log(h)|^2 h^2,$$

where the constant C_s does not depend on the mesh parameter h .

This theorem is proved in Rannacher & Vexler [81] using the techniques for estimating discrete Green functions developed in Frehse & Rannacher [51].

The obtained result holds also for the case of the parameter identification problem governed by more general elliptic equations of the form:

$$-\nabla \cdot (A(q)\nabla u + b_1(q)u) + b_2(q)u = f,$$

as well as by elliptic systems.

3.5. Numerical examples

In this section we discuss an example problem confirming the a priori error estimates from the previous section. The state equation is given by:

$$\begin{aligned} -\nabla \cdot (A(q)\nabla u) &= 2 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{3.5.72}$$

where Ω is the unit square. The matrix $A(q)$ is a function of the parameter $q = (q_1, q_2) \in Q = \mathbb{R}^2$ given by:

$$A(q) = \begin{pmatrix} q_1^2 & q_1 q_2 \\ q_1 q_2 & e^{q_2} \end{pmatrix}.$$

The parameters are estimated from the measurements of mean values along seven given lines:

$$\Gamma_i = (0, 1) \times \{0.125 i\}, \quad i = 1 \dots 7.$$

The components of the corresponding observation operator C have the following form:

$$C_i(v) = \int_{\Gamma_i} v ds, \quad i = 1 \dots 7.$$

The vector of measurements \bar{C} is given by:

$$\bar{C}_i = C_i(S(\hat{q}))(1 + \varepsilon_i), \quad i = 1 \dots 7,$$

where the reference parameter is $\hat{q} = (5, 6)$ and $\varepsilon = (\varepsilon_i)$ describes the data perturbation. We consider two cases:

$$\varepsilon = 0 \text{ and } \varepsilon \approx (0.12, -0.26, 0.29, -0.37, -0.49, 0.13, -0.04).$$

The parameter identification problem is discretized using bilinear finite elements on uniformly refined meshes. For the solution of the discretized problem we use Gauß-Newton method, described in the next chapter. The computations are done on a sequence of uniformly refined meshes such that maximal mesh size $h^{(k)}$ of the k -th mesh is given by

$$h^{(k)} = 2^{-k} h^{(0)}, \quad k = 0, 1, \dots 5.$$

This allows to estimate the order of convergence. The obtained results are listed in Table 3.5.1 and Table 3.5.2. For both cases the expected order of convergence is achieved.

TABLE 3.5.1. The error and the order of convergence with respect to the components of q without data perturbation.

h	$q_1 - q_{1,h}$	$q_2 - q_{2,h}$
2^{-3}	6.426e-1	4.525e-2
2^{-4}	2.455e-1	1.516e-2
2^{-5}	7.413e-2	4.347e-3
2^{-6}	1.971e-2	1.137e-3
2^{-7}	4.972e-3	2.853e-4
2^{-8}	1.192e-3	6.829e-5
order	2.024	2.029

TABLE 3.5.2. The error and the order of convergence with respect to the components of q with data perturbation.

h	$q_1 - q_{1,h}$	$q_2 - q_{2,h}$
2^{-3}	8.620e-2	1.583e-2
2^{-4}	2.099e-2	2.827e-3
2^{-5}	5.173e-3	5.800e-4
2^{-6}	1.294e-3	1.360e-4
2^{-7}	3.174e-4	3.122e-5
2^{-8}	7.544e-5	7.223e-6
order	2.050	2.117

CHAPTER 4

Optimization Algorithm

In this chapter we describe the optimization loop for the solution of the parameter identification problem discussed in the previous chapters. Later on, it is used on each mesh in an adaptive mesh refinement algorithm, see Chapter 5. In Section 4.1 we discuss a general class of Newton type methods for the solution of the parameter identification problem. Then, in Section 4.2 we analyze some realizations of the general optimization loop, namely full Newton method, Gauß-Newton method, quasi Newton methods with Broyden type updates and a special update method exploiting the structure of the parameter identification problem. Afterwards, in Section 4.3 we discuss trust region techniques for globalization of the convergence. All algorithms described here are applied to the reduced (unconstrained) formulation of the parameter identification problem (3.2.52). For a comparison of the algorithms based on the reduced and on the constrained formulation, see e.g. Hinze [64]. A generalized Gauß-Newton method for parameter identification problems in DAE systems is introduced by Bock [24, 25]. In the last section of this chapter we demonstrate the behavior of the presented algorithms for some numerical examples.

4.1. Newton type methods

In this section we describe a general Newton type optimization algorithm for the discretized parameter identification problem (3.2.49, 3.2.50). The Newton type methods are successfully used for parameter identification and optimization problems governed by systems of ODEs and DAEs, see e.g. Bock et. al. [25, 26], and also in the context of PDE constrained problems, see e.g. Becker [11], Hinze and Kunisch [65], and Haber, Ascher & Oldenburg [61].

The main idea by the Newton type methods is to compute a sequence of iterates $\{q_h^k\}$ starting with an initial guess q_h^0 and using a quadratic approximation of the cost functional for the recursive setting $q_h^{k+1} = q_h^k + \delta q_h$. The update δq_h is obtained using a symmetric approximation H_k of the Hessian $\nabla^2 j_h(q_h^k)$ as the solution of the system of linear equations:

$$H_k \delta q_h = -\nabla j_h(q_h^k). \quad (4.1.73)$$

The choice of the matrix H_k leads to different variants of the optimization algorithm, like Newton method, Gauß-Newton method, quasi Newton methods, see Section 4.2. In the following we give details for a typical optimization loop without specifying the choice of the matrix H_k .

For the computation of the gradient in the right-hand side of (4.1.73) we use the representation of the Jacobian matrix $G_h = c'_h(q_h^k)$ of the discrete reduced observation operator c_h from Proposition 3.2.2 and the fact, that

$$\nabla j_h(q_h^k) = G_h^*(c_h(q_h^k) - \bar{C}) + \alpha(q_h^k - \bar{q}). \quad (4.1.74)$$

The matrix G_h is also used by almost all variants of the optimization loop for the computation of the matrix H_k , see Section 4.2. In the stopping criterion we use a given tolerance ε .

Optimization Algorithm

(1) Choose an initial parameter $q_h^0 \in Q$ and set $k = 0$

(2) Compute $u_h^k \in V_h$, the solution of

$$a(u_h^k, q_h^k)(\phi_h) = f(\phi_h) \quad \forall \phi_h \in V_h$$

(3) Compute $w_{j,h}^k \in V_h$, $j = 1 \dots n_p$, the solutions of

$$a'_u(u_h^k, q_h^k)(w_{j,h}^k, \phi_h) = -a'_{q_j}(u_h^k, q_h^k)(1, \phi_h) \quad \forall \phi_h \in V_h$$

(4) Compute the matrix G_h by:

$$(G_h)_{ij} = C'_i(u_h^k)(w_{j,h}^k)$$

(5) Compute the residual $r_h \in Q$ as

$$r_h = G_h^*(\bar{C} - C(u_h^k)) + \alpha(\bar{q} - q_h^k)$$

(6) If $\|r_h\|_Q \leq \varepsilon$ quit

(7) Compute the matrix H_k

(8) Compute δq_h as solution of:

$$H_k \delta q_h = r_h$$

(9) Set $q_h^{k+1} = q_h^k + \delta q_h$

(10) Increment k and go to 2.

The local convergence theory for such algorithms is discussed in various publications, see e.g. Dennis & Schnabel [38] or Nocedal & Wright [78]. In Deuffhard, Engl & Scherzer [41] the convergence of iterative algorithms for least squares problem is analyzed. A general theorem about local convergence for Newton type methods is proved in Bock [25].

4.2. Realizations of the algorithm

In this section we shortly describe different realization of the algorithm presented in Section 4.1. We present Gauß-Newton method, full Newton method, quasi Newton method with Broyden type updates and a special Gauß-Newton update method, which exploits the structure of parameter identification problem and combines the advantages of the Gauß-Newton and Quasi-Newton methods.

4.2.1. GAUSS-NEWTON METHOD

The most widely used method for the solution of least squares problem is the Gauß-Newton method. The corresponding choice of the matrix H_k is

$$H_k = G_h^* G_h + \alpha I, \quad (4.2.75)$$

where $G_h = c'_h(q_h^k)$ is the Jacobian matrix of the discrete reduced observation operator c_h and is computed in Step 4 of the algorithm in Section 4.1. For this choice the solution of (4.1.73) can be interpreted as the solution of the linearized minimization problem

$$\text{Minimize } \frac{1}{2} \|c_h(q_h^k) + G_h \delta q_h - \bar{C}\|_Z^2 + \frac{\alpha}{2} \|q_h^k + \delta q_h - \bar{q}\|_Q^2. \quad (4.2.76)$$

For one Gauß-Newton step, the discrete state equation in Step 2 and n_p tangent problems in Step 3 have to be solved which originate from the same linear operator but with different right-hand sides. Due to the small dimension n_p of the parameter space Q the solution of (4.1.73) in Step 8 is uncritical.

This method has several advantages. First, the matrix H_k in (4.2.75) can be evaluated almost without extra cost using the already computed matrix G_h , which is required for the gradient evaluation in (4.1.74). Moreover, for the solution of (4.1.73) there is no need to calculate the matrix H_k explicitly, one can use QR or singular-value decompositions of the matrix G_h for solving the corresponding normal equation, see e.g. Stoer & Bulirsch [84]. Second, the matrix H_k is always positive semidefinite and there holds the following proposition:

PROPOSITION 4.2.1. *Let one of the following assumptions be valid:*

- (i) *The matrix G_h has full rank, i.e. $\text{Rank}(G_h) = n_p$,*
- (ii) *The regularization parameter α is positive.*

Then the Gauß-Newton matrix H_k , defined in (4.2.75), is positive definite.

PROOF. The proof is trivial. □

The positive definiteness of the matrix H_k implies that the corresponding update δq_h is always a descent direction of the cost function j_h . This allows more flexibility in the choice of globalization methods, i.e. line search is possible.

4.2.2. FULL NEWTON METHOD

The choice of the matrix H_k to be equal to the Hessian of the discrete reduced cost functional j_h , i.e.

$$H_k = \nabla^2 j_h(q_h^k)$$

leads to the full Newton method. For evaluation of the Hessian H_k one uses the representation given in Proposition 3.2.3, i.e.:

$$\nabla^2 j_h(q_h) = G_h^* G_h + M_h + \alpha I.$$

As before, the computation of the Jacobian G_h is required. The entries of the matrix M_h can be computed by a subtle evaluation of several second derivatives of the semi-linear form a in the directions of the solutions $w_{j,h} \in V_h$ of the tangent problems (3.2.55) and the solution $z_h \in V_h$ of the adjoint equation (3.2.58) as described in Proposition 3.2.3.

The Newton method produces a sequence of iterates $\{q_h^k\}$, which converges quadratically to the solution q_h provided the initial guess q_h^0 is in a sufficiently small neighborhood of the optimal parameter q_h , see e.g. Fletcher [50]. The Gauß-Newton method, described in the previous section, usually shows linear convergence.

In contrast to the Gauß-Newton method the matrices H_k are not necessary positive definite throughout the optimization loop, even if the positive definiteness of the Hessian is given in the solution q_h . This leads to the fact, that the produced sequence of iterates $\{q_h^k\}$ may converge to a local maximum or to a saddle point. One way to overcome this difficulty is the application of the Newton method combined with a trust region strategy, as described in Section 4.3.

Another difficulty by the Newton method is that the computation of the second derivatives of the semi-linear form a is in many cases very expensive. This is often a major disincentive to the use of the Newton method. In the following sections we describe two possibilities for improvement the local convergence properties of the Newton types algorithms without evaluation of the second derivatives.

4.2.3. QUASI-NEWTON METHODS

Quasi-Newton methods require only the computation of the gradient of the cost functional and construct the approximation H_k of the Hessian $\nabla^2 j_h(q_h^k)$ in an iterative way by an update formula. This construction is based on measuring the changes in gradients, which has to be in accordance with the new approximation H_{k+1} . This leads to a so called *secant equation* motivated by Taylor's formula:

$$H_{k+1} \delta q_h = \nabla j_h(q_h^{k+1}) - \nabla j_h(q_h^k). \quad (4.2.77)$$

However, the condition (4.2.77) does not determine the new approximation H_{k+1} . There several formulas for updating H_{k+1} . The most of them uses a correction of small rank, for instance of

rank two, i.e:

$$H_{k+1} = H_k + \beta x x^* + \gamma y y^*,$$

where $\beta, \gamma \in \mathbb{R}$ and $x, y \in \mathbb{R}^{np}$. In the following we describe the update formulas of Broyden class and a special update formula, which exploits the structure of parameter identification problem and combines the advantages of the Gauß-Newton and Quasi-Newton methods.

Broyden class

The most popular update formulas like BFGS and DFP are members of the Broyden class. The general update formula describing this family of methods has the following form:

$$H_{k+1} = H_k - \frac{x x^*}{x^* \delta q_h} + \frac{y y^*}{y^* \delta q_h} + \phi v v^*, \quad (4.2.78)$$

where

$$\begin{aligned} x &= H_k \delta q_h, \\ y &= \nabla j_h(q_h^{k+1}) - \nabla j_h(q_h^k), \end{aligned}$$

ϕ is a scalar parameter and v is given by:

$$v = \frac{y}{y^* \delta q_h} - \frac{x}{x^* \delta q_h}.$$

The BFGS formula is obtained from (4.2.78) by setting $\phi = 0$ and the DFP formula by setting $\phi = 1$. The application of update formulas of Broyden class leads to local superlinear convergence. For detailed analysis of these formulas see e.g. Fletcher [50] or Nocedal & Wright [78].

Gauß-Newton-Update method

For the solution of parameter identification problems one can derive an update formula exploiting the structure of the Hessian $\nabla^2 j_h(q_h^k)$. By using the algorithm described in Section 4.1 the information of the Gauß-Newton part of the Hessian is evaluable. Therefore, it is natural to consider the following approximation H_k :

$$H_k = G_h^* G_h + \alpha I + \widehat{M}_k,$$

where \widehat{M}_k is an approximation of the matrix M_h from (3.2.57) and is given by an update formula. By construction of an update formula for \widehat{M}_k one has to ensure that the secant condition (4.2.77) holds for the matrix H_k . Such an update formula was suggested in Dennis, Gay & Welsch [39]. There, the matrix \widehat{M}_k is recursively updated starting with $\widehat{M}_0 = 0$:

$$\widehat{M}_{k+1} = \widehat{M}_k + \frac{1}{y^* \delta q} (x y^* + y x^*) - \frac{x^* \delta q}{(y^* \delta q)^2} y y^*,$$

where

$$\begin{aligned} y &= G_{k+1}^* r_{k+1} - G_k^* r_k, \\ x &= (G_{k+1}^* - G_k^*) r_{k+1} - \widehat{M}_k \delta q. \end{aligned}$$

For derivation and analysis of this update formula, see also Dennis & Schnabel [38].

4.3. Trust region method

It is well known, see e.g. Fletcher [50], that the convergence of the algorithms described so far is ensured, only if the initial guess q_h^0 is in a sufficiently small neighborhood of the optimal parameter q_h . There are two main possibilities to overcome this difficulty and to improve the global convergence: line search and trust region methods, see e.g. Nocedal & Wright [78]. The line search method requires positive definiteness of the matrices H_k throughout the optimization algorithm. This is often too restrictive for the kind of parameter identification problems we consider, see examples in Section 4.4. Therefore, we apply trust region techniques for improving the global convergence. In the following we describe the algorithm we use.

If the matrix H_k is positive definite, the computation of $\delta q_h \in Q$ in (4.1.73) can be interpreted as the solution of a minimization problem:

$$\text{Minimize } m_k(\delta q_h) := j_h(q_h^k) + \nabla j_h(q_h^k)^* \delta q_h + \frac{1}{2} \delta q^* H_k \delta q, \quad \delta q \in Q, \quad (4.3.79)$$

where $\nabla j_h(q_h^k)$ is given by (4.1.74). The cost functional m_k of (4.3.79) is the so called *local model function*, whose behavior near the current point q_h^k is similar to that of the actual cost functional j_h defined in (3.2.52). However, the local model function m_k may not be a good approximation of j_h for large δq_h . Therefore, we restrict the search for a minimizer of m_k to a ball (*trust region*) around q_h^k . In other words, we replace the problem (4.3.79) by the following constrained optimization problem:

$$\text{Minimize } m_k(\delta q_h), \quad \text{subject to } \|\delta q_h\|_Q \leq \Delta_k, \quad (4.3.80)$$

with a trust region radius Δ_k to be determined iteratively.

For the convergence properties of the trust region method, the strategy for choosing the trust region radius Δ_k is crucial. Following the standard approach, see e.g. Conn, Gould & Toint [36], we base this choice on the agreement between the model function m_k and the cost functional j_h at the previous iteration. For the increment δq_h , we define the ratio

$$\rho_k = \frac{j_h(q_h^k) - j_h(q_h^k + \delta q_h)}{m_k(0) - m_k(\delta q_h)}, \quad (4.3.81)$$

and use it as an indicator of the quality of the local model m_k . The numerator of (4.3.81) is called the *actual reduction*, and the denominator is the *predicted reduction*. If this ratio is close to 1, there is a good agreement between the model m_k and the cost functional j_h for the current step. As a consequence, the trust region is expanded for the next iteration. Otherwise, we do not alter the trust region or shrink it, depending on the distance $|\rho_k - 1|$. We note, that due to the fact, that δq_h is a minimizer of (4.3.80), the predicted reduction is always nonnegative. Therefore, if ρ_k is negative, the new value $j_h(q_h^k + \delta q_h)$ is greater than the current value $j_h(q_h^k)$ and the step is rejected. Let the initial and the maximal trust region radius Δ_0 and $\bar{\Delta}$ be given,

then a typical procedure for determining the new trust region radius has the following form:

$$\Delta_{k+1} = NTR(\Delta_k, \rho_k, \delta q_h) = \begin{cases} \frac{1}{4}\|\delta q_h\|_Q, & \text{if } \rho_k < \frac{1}{4}, \\ \Delta_k, & \text{if } \frac{1}{4} < \rho_k < \frac{3}{4}, \\ \min(2\Delta_k, \bar{\Delta}), & \text{if } \rho_k > \frac{3}{4}. \end{cases} \quad (4.3.82)$$

For a given $\epsilon > 0$ and $\sigma \in (0, \frac{1}{4})$ the trust region algorithm is as follows:

Trust Region Algorithm

- (1) Choose an initial parameter $q_h^0 \in Q$ and set $k = 0$
- (2) Compute $r_h = \nabla j_h(q_h^k)$
- (3) Compute the matrix H_k
- (4) If $\|r_h\|_Q \leq \epsilon$ quit
- (5) Compute δq_h by solving (4.3.80)
- (6) Compute ρ_k by (4.3.81)
- (7) Compute $\Delta_{k+1} = NTR(\Delta_k, \rho_k, \delta q_h)$, see (4.3.82)
- (8) If $\rho_k > \sigma$ set $q_h^{k+1} = q_h^k + \delta q_h$
- (9) Increment k and go to 2.

REMARK 4.3.1. The residual r_h is computed in the same way as in the algorithm in Section 4.1. The matrix H_k is obtained by one of the methods described in Section 4.2.

REMARK 4.3.2. For the stability of the algorithm the current step is rejected not only if $\rho_k < 0$ but also if $\rho_k \leq \sigma$, with a small positive number σ .

Due to the compactness of the feasible set described by the condition $\|\delta q_h\|_Q \leq \Delta_k$, the problem (4.3.80) possess always a solution independently of the definiteness of the matrix H_k . The procedure we use for solving (4.3.80) in Step 5 is based on the following proposition characterizing the solution of (4.3.80).

PROPOSITION 4.3.3. *The vector δq_h is a global solution of the trust region problem (4.3.80) if and only if there is a scalar $\lambda \geq 0$ such that the following conditions are fulfilled:*

(i) $(H_k + \lambda I) \delta q_h = -\nabla j_h(q_h^k),$

(ii) $\lambda(\Delta_k - \|\delta q_h\|) = 0,$

$$(iii) \quad \|\delta q_h\| \leq \Delta_k,$$

(iv) $(H_k + \lambda I)$ is positive semidefinite.

PROOF. For proof see e.g. Nocedal & Wright [78]. \square

For the solution of (4.3.80) we compute an eigenvalue decomposition of the symmetric matrix H_k , i.e.

$$H_k = T D T^*,$$

where T is an orthogonal matrix and D is a diagonal matrix

$$D = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{n_p})$$

with $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_{n_p}$ eigenvalues of H_k . Due to the small dimension n_p of the parameter space Q the effort for this decomposition is uncritical.

If the matrix H_k is positive definite (i.e. $\lambda_1 > 0$) and it holds

$$\|H_k^{-1} \nabla j_h(q_h^k)\|_Q \leq \Delta_k,$$

we set $\delta q_h = -H_k^{-1} \nabla j_h(q_h^k)$. Otherwise, the solution δq_h is searched on the boundary of the feasible set, i.e. such that $\|\delta q\|_Q = \Delta_k$. Using Proposition 4.3.3 we have:

$$\delta q_h = -T (D + \lambda I)^{-1} T^* \nabla j_h(q_h^k).$$

Hence, denoting $g = T^* \nabla j_h(q_h^k)$ we obtain:

$$\|\delta q_h\|_Q^2 = \|T (D + \lambda I)^{-1} g\|_Q^2 = g^* ((D + \lambda I)^{-1})^2 g = \sum_{j=1}^{n_p} \frac{g_j^2}{(\lambda_j + \lambda)^2},$$

where g_j is the j th component of $g \in Q$. Therefore, we need only to solve a one-dimensional root-finding problem

$$\sum_{j=1}^{n_p} \frac{g_j^2}{(\lambda_j + \lambda)^2} = \Delta_k^2,$$

which is done by using a standard Newton algorithm.

The following theorem (see [78]) ensures the global convergence of the presented method:

THEOREM 4.3.1. *Suppose that $\|H_k\| \leq \beta$ for some constant β and the level set*

$$\mathcal{N} = \{q \mid j_h(q) \leq j_h(q_h^0)\}$$

is bounded. Then there holds for the iterates q_h^k obtained by the algorithm described above:

$$\lim_{k \rightarrow \infty} \nabla j_h(q_h^k) = 0.$$

4.4. Numerical examples

In this section we demonstrate the behavior of the algorithms described above for simple numerical examples. We consider a convection-diffusion equation with unknown constant transport direction (q_1, q_2) in the unit square $\Omega = (0, 1)^2$:

$$\begin{aligned} -\Delta u + q_1 u_x + q_2 u_y &= 2 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \quad (4.4.83)$$

The parameter q is estimated using measurements given by the values of the state variable at five different points:

$$\begin{aligned} \xi_1 &= (0.25, 0.5), & \xi_2 &= (0.5, 0.25), \\ \xi_3 &= (0.75, 0.5), & \xi_4 &= (0.5, 0.75), & \xi_5 &= (0.5, 0.5). \end{aligned} \quad (4.4.84)$$

The components of the corresponding observation operator C have the following form:

$$C_i(v) = v(\xi_i), \quad (4.4.85)$$

and the parameter identification problem is formulated as follows: For $(u, q) \in V \times Q$ with $V = H_0^1(\Omega)$ and $Q = \mathbb{R}^2$

$$\text{Minimize } \frac{1}{2} \sum_{i=1}^5 (u(\xi_i) - \bar{C}_i)^2 \quad (4.4.86)$$

under the constraint (4.4.83). The components of the measurement vector $\bar{C} \in Z = \mathbb{R}^5$ are given by

$$\bar{C}_i = S(\hat{q})(\xi_i)(1 + \varepsilon_i),$$

where $\hat{q} = (8, 8)$ and

$$|\varepsilon_i| \leq \delta, \quad i = 1, 2, 3, 4, 5.$$

For our numerical results we use pseudo-random perturbations ε_i with different perturbations levels δ . We consider three cases: $\delta = 0.1, 0.3, 0.5$ with corresponding perturbations:

$$\begin{aligned} \varepsilon^1 &= (0.043, -0.008, 0.068, -0.084, -0.041), \\ \varepsilon^2 &= (0.092, -0.182, -0.285, 0.039, -0.146), \\ \varepsilon^3 &= (0.121, -0.255, 0.285, -0.370, -0.492). \end{aligned} \quad (4.4.87)$$

The solutions of the corresponding problems are:

$$\begin{aligned} q^1 &\approx (8.52229, 7.06701), \\ q^2 &\approx (7.25298, 11.4838), \\ q^3 &\approx (11.1831, 4.55648). \end{aligned} \quad (4.4.88)$$

In Figure 4.4 we see the isolines of the reduced observation operator $j(q)$ depending on q for the third case. We note, that the initial guess for optimization $q_0 = (0, 0)$ lies in a neighborhood

of a local maximum of the cost functional j and the corresponding Hessian matrix $\nabla^2 j(q_0)$ is negative definite with eigenvalues $(\lambda_1, \lambda_2) \approx (-4.2e-4, -3.9e-4)$.

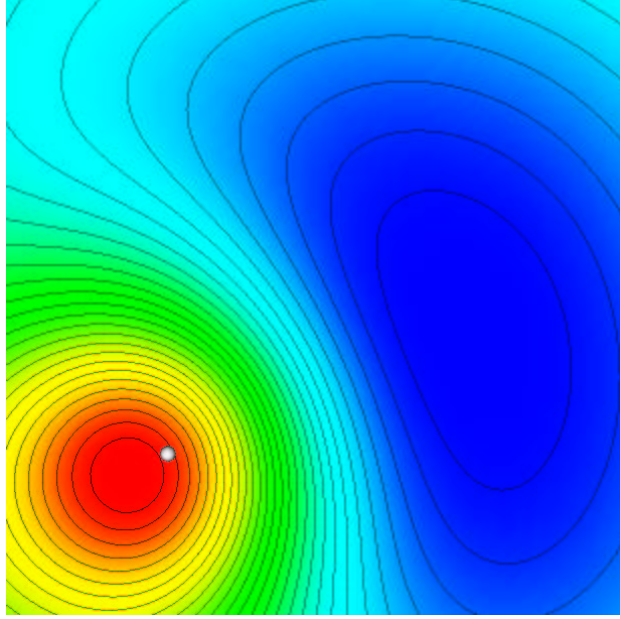


FIGURE 4.4.1. The isolines of the reduced observation operator $j(q)$ depending on q . The initial guess for optimization $q_0 = (0, 0)$ is marked by a ball.

For these problems we compare different optimization algorithms described in Section 4.2. For all algorithms here we use the same trust region techniques described in Section 4.3. In Table 4.4.1 we compare the number of iterations for different optimization algorithms (Newton method, Gauß-Newton method, BFGS method, and the Gauss-Newton-Update method) in the first case ($\delta = 0.1$). The comparison is done on three uniform refined meshes with 256, 1024 and 4096 nodes. The corresponding results for $\delta = 0.3$ and $\delta = 0.5$ are given in Table 4.4.2 and Table 4.4.3.

TABLE 4.4.1. Number of iterations of different optimization algorithms for $\delta = 0.1$

N	Gauß-Newton	Newton	Gauß-Newton-Update	BFGS
256	11	9	10	24
1024	5	3	5	20
4096	5	3	5	11

TABLE 4.4.2. Number of iterations of different optimization algorithms for $\delta = 0.3$

N	Gauß-Newton	Newton	Gauß-Newton-Update	BFGS
256	18	10	10	28
1024	8	3	6	21
4096	7	3	6	17

TABLE 4.4.3. Number of iterations of different optimization algorithms for $\delta = 0.5$

N	Gauß-Newton	Newton	Gauß-Newton-Update	BFGS
256	60	9	11	27
1024	49	3	7	18
4096	45	3	6	17

CHAPTER 5

A Posteriori Error Analysis

This chapter is devoted to the a posteriori error estimation. Our goal is the development of an a posteriori error estimator for the error in the parameters. Its purpose is to guide an adaptive mesh refinement algorithm. Furthermore, the estimator is used to assess the accuracy of the computed parameters.

The concepts of adaptivity and a posteriori error estimation are now commonly accepted for the numerical solution of partial differential equations. The research in this field was initiated by the pioneering work of Babuška & Rheinboldt [4, 5] and further developed e.g. by Bank & Weiser [8], Babuška & Miller [6] and Ainsworth & Oden [1]. The use of “duality techniques” for a posteriori error estimation in finite element methods is first systematically pursued by Eriksson & Johnson [43, 44]. In Becker & Rannacher [18, 19] this approach is further developed into a computation-based feedback method for error control and mesh optimization with respect to a target functional representing physical quantity of interest.

However, to our knowledge, there are only few published results on adaptive finite elements for optimization problems, see Becker [11], Becker, Kapp & Rannacher [20] and Liu & Yan [73]. In Bangerth [7] adaptivity is used for estimating distributed parameters of elliptic partial differential equations. Our approach to a posteriori error estimation for parameter identification problems presented below, is developed in Becker & Vexler [21, 22]. It is based on the optimal control approach to a posteriori error estimation from Becker & Rannacher [18, 19]. However, a direct application of the techniques described in [18, 19] leads to an estimator which controls the error in the cost functional J (2.1.2), see [7]. In general, such an estimator does not provide useful error bounds for the parameters, in contrast to the approach presented here.

In order to measure the error in the parameters, we introduce an error functional $E : Q \rightarrow \mathbb{R}$. The use of the error functional E allows one to weight the relative importance of the different parameters. We prove the following error representation:

$$E(q) - E(q_h) = \eta_h + \mathcal{R}, \tag{5.0.89}$$

where $E(q) - E(q_h)$ describe the error between the solution q of the continuous parameter identification problem and the solution q_h of the discrete one. Further, η_h denotes the a posteriori error estimator and \mathcal{R} consists of remainder terms, which may usually be neglected; see the

discussion in Section 5.2.

The error estimator η_h is used within an adaptive algorithm for error control and mesh refinement, sketched below. Such an algorithm generates a sequence of locally refined meshes and corresponding finite element spaces until the estimated error with respect to E is below a given tolerance TOL . For the following iteration, we use a mesh refinement procedure, which adaptively refines a given regular mesh to obtain a new regular mesh for the next iteration. It is guided by information based on the localization of the error estimator η_h to cell-wise (or node-wise) contributions, see Section 5.4.

Adaptive Mesh Refinement Algorithm

- (1) Choose an initial mesh \mathcal{T}_{h_0} and set $k = 0$
- (2) Construct the finite element space V_{h_k}
- (3) Compute $u_{h_k} \in V_{h_k}, q_{h_k} \in Q$ by solving (3.2.49, 3.2.50)
- (4) Evaluate the a posteriori error estimator η_{h_k}
- (5) If $\eta_{h_k} \leq TOL$ quit
- (6) Refine $\mathcal{T}_{h_k} \rightarrow \mathcal{T}_{h_{k+1}}$ using information from η_{h_k}
- (7) Increment k and go to 2.

REMARK 5.0.1. In step 3, the discretized parameter identification problem (3.2.49, 3.2.50) is solved on the current mesh \mathcal{T}_{h_k} , see Chapter 4. As initial data, we use the values from the computation on the previous mesh. This allows us to avoid unnecessary iterations of the optimization loop on fine meshes.

This chapter is organized as follows: In Section 5.1 we discuss the a posteriori error estimation for a simple linear example in order to illustrate our approach. In Section 5.2 we derive the error estimator for the error in parameters in the general nonlinear case without assuming a perfect match, i.e without assuming $C(u) = \bar{C}$. We present two versions of the a posteriori error estimator. For the first one the knowledge of the Hessian of the reduced cost functional j_h is required, which is computed in the last step of the optimization loop by using Newton method, as described in Section 4.2.2. Further, a simplified version of the error estimator is derived, which does not require the use of the Newton method, but leads to an additional remainder term, see the discussion in Section 5.2. An extension to the class of problems, where the quantity of interest (error functional E) depends on the state variable as well as on the parameters, is discussed in Section 5.3. In Section 5.4 we present the techniques for the localization of the error estimator to node-wise (or cell-wise) contributions. Numerical examples illustrating the

behavior of the adaptive mesh refinement algorithm based on our error estimator are presented in Section 5.5. We consider linear and nonlinear elliptic equations where the parameters enter in different ways.

5.1. An introductory example

We discuss the question of a posteriori error estimation for a simple linear example with one parameter and one observation, i.e. $Q = Z = \mathbb{R}$. The state equation, the observation operator and the dependency of the state variable on the parameter are all linear. The regularization parameter α is zero. The problem is formulated as follows: for a positive weight function $\omega \in L^2(\Omega)$ and a number \bar{C} we minimize the least squares functional

$$J(u) = \frac{1}{2} \left| \int_{\Omega} \omega u \, dx - \bar{C} \right|^2 \quad (5.1.90)$$

under the constraint

$$\begin{aligned} -\Delta u &= qf && \text{in } \Omega = (0, 1)^2, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned} \quad (5.1.91)$$

with positive $f \in L^2(\Omega)$ and $u \in V = H_0^1(\Omega)$. A standard weak formulation of the state equation (5.1.91) is given by:

$$(\nabla u, \nabla \phi) = (qf, \phi) \quad \forall \phi \in V. \quad (5.1.92)$$

The meaning of this example problem is, that we adjust the scaling of the load in such a way, that the weighted mean of the solution u attains a prescribed value.

Here, the solution operator is given by

$$u = S(q) = qu_1, \quad (5.1.93)$$

where u_1 is the solution of the state equation (5.1.91) for $q = 1$, i.e.

$$\begin{aligned} -\Delta u_1 &= f && \text{in } \Omega \\ u_1 &= 0 && \text{on } \partial\Omega. \end{aligned} \quad (5.1.94)$$

Therefore the optimal parameter is simply given by:

$$q = \bar{C}\mu, \quad \mu := \frac{1}{\int_{\Omega} \omega u_1 \, dx}. \quad (5.1.95)$$

By virtue of the maximum principle (see e.g. Hellwig [62]), μ is well defined. Similarly, the optimal parameter for the finite element discretization of the problem (5.1.90, 5.1.91), see Chapter 3, is given by

$$q_h = \bar{C}\mu_h, \quad \mu_h := \frac{1}{\int_{\Omega} \omega u_{1h} \, dx}, \quad (5.1.96)$$

where $u_{1h} \in V_h$ is the solution of the discrete analog of (5.1.94), i.e.:

$$(\nabla u_{1h}, \nabla \phi_h) = (f, \phi_h) \quad \forall \phi_h \in V_h.$$

Now, the fundamental idea is to introduce the adjoint equation:

$$\begin{aligned} -\Delta y &= -\mu\omega && \text{in } \Omega \\ y &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{5.1.97}$$

Using the equation (5.1.97) we prove the following error representation:

PROPOSITION 5.1.1. *For the finite element discretization of the simple example (5.1.90, 5.1.91), we have the following error representation:*

$$q - q_h = \rho(y - i_h y), \tag{5.1.98}$$

where y is the solution of the adjoint equation (5.1.97) and ρ is a residual of the state equation (5.1.92) defined by:

$$\rho(\phi) := (q_h f, \phi) - (\nabla u_h, \nabla \phi).$$

PROOF. We have, using the definitions of q , μ and (5.1.97):

$$\begin{aligned} q - q_h &= \bar{C}\mu - q_h \mu \int_{\Omega} \omega u_1 dx \\ &= \mu \int_{\Omega} \omega u_h dx + q_h (\nabla u_1, \nabla y) \\ &= -(\nabla u_h, \nabla y) + (q_h f, \nabla y). \end{aligned}$$

We complete the proof using the Galerkin orthogonality relation, i.e. the fact that:

$$(q_h f, \phi_h) - (\nabla u_h, \nabla \phi_h) = 0 \quad \forall \phi_h \in V_h.$$

□

REMARK 5.1.2. A direct application of the approach in Becker, Rannacher & Kapp [20] would lead to consider the following adjoint equation:

$$\begin{aligned} -\Delta z &= -(\bar{C} - \int_{\Omega} \omega u dx)\omega && \text{in } \Omega \\ z &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{5.1.99}$$

The right hand side in (5.1.99) is the derivative of the least squares functional. This implies $z = 0$ in our case, and therefore, the resulting error bound is useless.

REMARK 5.1.3. The simple form and derivation of error representation (5.1.98) are due to the structure of the problem: beside the linearity and $n_m = n_p = 1$, the fact that the observations are matched, i.e. that the least squares residual $C(u) - \bar{C}$ vanishes, is crucial. In the general setting later on, we use an approach based on a special Lagrange function.

REMARK 5.1.4. The quantity μ captures the stability of the problem, since $\mu = 1/c'(q)$ with the reduced functional c defined in (2.3.23). It can also be seen that $\mu = H^{-1}G^*$ with $H = \nabla^2 j(q)$, which is the correct expression in the general setting. The same term appears in the sensitivity analysis, cf. Theorem 2.4.2 in Section 2.4.

REMARK 5.1.5. For evaluation of the error representation (5.1.98), the local interpolation error $y - i_h y$ has to be approximated. Later on, we use interpolation of the computed bilinear finite element solution y_h on the space of biquadratic finite elements on patches of cells, see Section 5.4.

5.2. A posteriori error estimation for the error in parameters

Here, we derive our a posteriori error estimator for the error in the parameters in the general setting. More precisely, we estimate the error with respect to a given error functional $E : Q \rightarrow \mathbb{R}$. The functional E is supposed to be three times continuously differentiable and we define its gradient by identification in the usual way:

$$\langle \nabla E(q), \delta q \rangle_Q = E'(q)(\delta q) \quad \forall \delta q \in Q.$$

Our goal is to prove the following error representation:

$$E(q) - E(q_h) = \eta_h + R,$$

where η_h denotes the a posteriori error estimator to be developed and R is a cubic remainder term due to linearization. The precise result is given in Theorem 5.2.1. The evaluation of this error estimator requires knowledge of certain second derivatives of the reduced observation operator c , which are computed in the last step of the optimization algorithm by using Newton method. Further, we derive a simplified version of the error estimator, which is cheaper to evaluate and does not require the use of Newton method; here additional remainder terms related to the least squares residual $R^{LS}(u) = \bar{C} - C(u)$ appear, see Theorem 5.2.2.

For the derivation of our a posteriori error estimator we make the following preparations: We recall that the optimality conditions for the parameter identification problem (2.1.1, 2.1.2) are based on the Lagrange function \mathcal{L} introduced in Section 2.3,

$$\mathcal{L}(u, q, z) = \frac{1}{2} \|C(u) - \bar{C}\|_Z^2 + \frac{\alpha}{2} \|q - \bar{q}\|_Q^2 + f(z) - a(u, q)(z),$$

for $u \in V$, $q \in Q$ and $z \in V$. The expression of the derivatives of \mathcal{L} is given using the abbreviation:

$$\begin{aligned}\rho_u(\xi)(\phi) &:= f(\phi) - a(u, q)(\phi), \\ \rho_z(\xi)(\psi) &:= -\langle C'(u)(\psi), R^{LS}(u) \rangle_Z - a'_u(u, q)(\psi, z), \\ \rho_q(\xi)(\sigma) &:= \alpha \langle q - \bar{q}, \sigma \rangle_Q - a'_q(u, q)(\sigma, z),\end{aligned}\tag{5.2.100}$$

where $\phi, \psi \in V$ and $\sigma \in Q$ are test functions; and we have with $\delta\xi = (\delta u, \delta q, \delta z)$:

$$\mathcal{L}'(\xi)(\delta\xi) = \rho_u(\xi)(\delta z) + \rho_z(\xi)(\delta u) + \rho_q(\xi)(\delta q).$$

For the error functional E , we introduce the following Lagrangian \mathcal{M} :

$$\mathcal{M}(\xi, \chi) = E(q) + \mathcal{L}'(\xi)(\chi),\tag{5.2.101}$$

where we have introduced an additional set of variables $\chi = (v, p, y) \in V \times Q \times V$. Similar to Becker & Rannacher [19], we have an error representation expressed in the following proposition.

PROPOSITION 5.2.1. *Let $x = (\xi, \chi) \in X = (V \times Q \times V)^2$ be a stationary point of \mathcal{M} , i.e.*

$$\mathcal{M}'(x)(\delta x) = 0 \quad \forall \delta x \in X.\tag{5.2.102}$$

Further let $X_h = (V_h \times Q \times V_h)^2 \subset X$ be a subspace and $x_h = (\xi_h, \chi_h) \in X_h$ the corresponding Galerkin solution:

$$\mathcal{M}'(x_h)(\delta x_h) = 0 \quad \forall \delta x_h \in X_h.\tag{5.2.103}$$

Then, there holds

$$E(q) - E(q_h) = \frac{1}{2} \mathcal{M}'(x_h)(x - \hat{x}_h) + R,\tag{5.2.104}$$

where $\hat{x}_h \in X_h$ is arbitrary and the remainder term R is given by:

$$R = \frac{1}{2} \int_0^1 \mathcal{M}'''(x_h + s e)(e, e, e) s(s-1) ds,\tag{5.2.105}$$

with $e = x - x_h$.

PROOF. We note, that ξ is a stationary point of \mathcal{L} , i.e.

$$\mathcal{L}'(\xi)(\delta\xi) = 0 \quad \forall \delta\xi \in V \times Q \times V\tag{5.2.106}$$

and ξ_h is the corresponding Galerkin solution

$$\mathcal{L}'(\xi_h)(\delta\xi_h) = 0 \quad \forall \delta\xi_h \in V_h \times Q \times V_h.\tag{5.2.107}$$

Therefore, we obtain:

$$E(q) - E(q_h) = \mathcal{M}(x) - \mathcal{M}(x_h).\tag{5.2.108}$$

We rewrite the right hand side of (5.2.108) as follows:

$$\mathcal{M}(x) - \mathcal{M}(x_h) = \int_0^1 \mathcal{M}'(x_h + se)(e)ds, \quad (5.2.109)$$

approximate the integral by the trapezoidal rule and obtain:

$$\mathcal{M}(x) - \mathcal{M}(x_h) = \frac{1}{2}\mathcal{M}'(x)(e) + \frac{1}{2}\mathcal{M}'(x_h)(e) + R, \quad (5.2.110)$$

where the remainder term R is defined in (5.2.105). The term $\mathcal{M}'(x)(e)$ vanishes, and due to Galerkin orthogonality the term $\mathcal{M}'(x_h)(e)$ can be replaced by $\mathcal{M}'(x_h)(x - \widehat{x}_h)$ with \widehat{x}_h arbitrarily chosen. This completes the proof. \square

Next, we consider the derivative \mathcal{M}' . First, using (5.2.101) and the splitting of variables $x = (\xi, \chi)$, $\delta x = (\delta\xi, \delta\chi)$, we have:

$$\mathcal{M}'(x)(\delta x) = \mathcal{L}'(\xi)(\delta\chi) + E'(q)(\delta q) + \mathcal{L}''(\xi)(\delta\xi, \chi). \quad (5.2.111)$$

Therefore, the stationarity condition (5.2.102) splits into the following two systems of equations. First, the original variables ξ are determined by the stationarity of \mathcal{L} , which corresponds to variations $\delta\chi$. Second, the auxiliary variables χ are solutions of the equations corresponding to variations $\delta\xi$. In order to rewrite the last terms, we introduce some additional residual functionals:

$$\begin{aligned} \rho_v(x)(\phi) &:= -a'_q(u, q)(p, \phi) - a'_u(u, q)(v, \phi), \\ \rho_y(x)(\psi) &:= \langle C'(u)(v), C'(u)(\psi) \rangle_Z - \langle C''(u)(\psi, v), R^{LS}(u) \rangle_Z \\ &\quad - a''_{uu}(u, q)(\psi, v, z) - a''_{uq}(u, q)(\psi, p, z) - a'_u(u, q)(\psi, y), \\ \rho_p(x)(\sigma) &:= E'(q)(\sigma) - a''_{uq}(u, q)(v, \sigma, z) - a''_{qq}(u, q)(\sigma, p, z) \\ &\quad - a'_q(u, q)(\sigma, y) + \alpha \langle p, \sigma \rangle_Q. \end{aligned} \quad (5.2.112)$$

With these notations, we have:

$$\begin{aligned} \mathcal{M}'(x)(\delta x) &= \{\rho_u(\xi)(\delta y) + \rho_z(\xi)(\delta v) + \rho_q(\xi)(\delta p)\} \\ &\quad + \{\rho_v(x)(\delta z) + \rho_y(x)(\delta u) + \rho_p(x)(\delta q)\}. \end{aligned} \quad (5.2.113)$$

At first glance, it seems as if application of Proposition 5.2.1 requires a huge coupled system to be solved. That this is not the case, is shown in Proposition 5.2.2. It turns out that the main afford in computation of the auxiliary solution (v, p, y) is already done in the optimization loop based on Newton method as described in Chapter 4.

PROPOSITION 5.2.2. *Let (u, q, z) be a stationary point of \mathcal{L} (2.3.37) and let $H = (H_{jk})$ be the reduced Hessian defined by*

$$H = G^*G + \alpha I + M, \quad (5.2.114)$$

where $M = -\sum_{i=1}^{n_m} c_i''(q)R_i^{LS}$. Further, let $\{w_j\}_{1 \leq j \leq n_p}$ be the tangent solutions defined in (2.3.30). Then, the auxiliary solution (v, p, y) is given by:

$$Hp = -\nabla E, \quad v = \sum_{j=1}^{n_p} w_j p_j, \quad (5.2.115)$$

and y is the solution of the following equation:

$$\begin{aligned} a'_u(u, q)(\phi, y) &= \langle Gp, C'(u)(\phi) \rangle_Z - \langle C''(u)(\phi, v), R^{LS}(u) \rangle_Z \\ &\quad - a''_{uu}(u, q)(\phi, v, z) - a''_{uq}(u, q)(\phi, p, z) \quad \forall \phi \in V. \end{aligned} \quad (5.2.116)$$

PROOF. (u, q, z) is a stationary point of \mathcal{L} and consequently the residuals $\rho_u(x)$, $\rho_z(x)$ and $\rho_q(x)$ vanish. Due to the construction of v in (5.2.115) we obtain:

$$\begin{aligned} \rho_v(x)(\phi) &= -\sum_{j=1}^{n_p} p_j a'_u(u, q)(w_j, \phi) - a'_q(u, q)(p, \phi) \\ &= -\sum_{j=1}^{n_p} p_j (a'_u(u, q)(w_j, \phi) + a'_q(u, q)(1, \phi)). \end{aligned}$$

This sum vanishes because of the definition of w_j (2.3.30). Moreover, using Proposition 2.3.1 we obtain:

$$C'_i(u)(v) = \sum_{j=1}^{n_p} p_j C'_i(u)(w_j) = \sum_{j=1}^{n_p} G_{ij} p_j = (Gp)_i.$$

Therefore, $C'(u)(v) = Gp$ and $\rho_y(x)$ vanishes too due to the definition of y (5.2.116). Finally, in order to see, that $\rho_p(x) = 0$, we obtain using (5.2.115), (5.2.116) and the representations of G and M given by Proposition 2.3.1 and Proposition 2.3.2:

$$\begin{aligned} E'_{q_j}(q) &= -\sum_{k=1}^{n_m} G_{kj}(Gp)_k - \sum_{k=1}^{n_m} M_{jk} p_k - \alpha p_j \\ &= -\langle Gp, C'(u)(w_j) \rangle_Z + \langle C''(u)(w_j, v), R^{LS}(u) \rangle_Z \\ &\quad + a''_{uu}(u, q)(w_j, v, z) + a''_{uq}(u, q)(w_j, p, z) + a''_{uq_j}(u, q)(v, 1, z) \\ &\quad + a''_{qq_j}(u, q)(p, 1, z) - \alpha p_j \\ &= a''_{uq_j}(u, q)(v, 1, z) + a''_{qq_j}(u, q)(p, 1, z) - a'_u(u, q)(w_j, y) - \alpha p_j \\ &= a'_q(u, q)(1, y) + a''_{uq_j}(u, q)(v, 1, z) + a''_{qq_j}(u, q)(p, 1, z) - \alpha p_j. \end{aligned}$$

This completes the proof. \square

REMARK 5.2.3. The construction of the corresponding discrete set of auxiliary variables $\chi_h = (v_h, p_h, y_h) \in V_h \times Q \times V_h$ is done in the same way.

From Propositions 5.2.1 and 5.2.2 we obtain the following result using a suitable interpolation operator $i_h : V \rightarrow V_h$, see, e.g., Clément [35].

THEOREM 5.2.1. *Let $\xi = (u, q, z)$ be the solution of the parameter identification problem and let $\chi = (v, p, y)$ be defined by equations (5.2.115) and (5.2.116). Further we denote by $\xi_h = (u_h, q_h, z_h)$ and $\chi_h = (v_h, p_h, y_h)$ the corresponding Galerkin solutions, and set $x_h = (\xi_h, \chi_h)$. Then, the following error representation holds:*

$$E(q) - E(q_h) = \frac{1}{2} \{ \rho_u(\xi_h)(y - i_h y) + \rho_z(\xi_h)(v - i_h v) \} \quad (5.2.117)$$

$$+ \frac{1}{2} \{ \rho_v(x_h)(z - i_h z) + \rho_y(x_h)(u - i_h u) \} + R, \quad (5.2.118)$$

where R is the cubic remainder term defined in (5.2.105).

PROOF. We use Proposition 5.2.1 and set $\hat{x}_h = (i_h u, q, i_h z, i_h v, p, i_h y)$. This complete the proof. \square

Theorem 5.2.1 gives a satisfactory result in the following sense: with the solution of only one auxiliary problem, i.e. the discrete version of (5.2.116), we obtain an error representation with a cubic remainder term. For a discussion of the estimation of such remainder terms from linearization, see Vexler [86]. Moreover, no assumption on the smallness of the least squares residual $R^{LS}(u)$ is necessary. However, the Newton matrix H , and therefore computation of the adjoint solution z are required.

In the following, we provide a simpler error representation which can be used when the Gauß-Newton algorithm is employed in the optimization loop. For this purpose, we split the derivative \mathcal{M}' into two parts:

$$\mathcal{M}'(x)(\delta x) = D(x)(\delta x) + T(x)(\delta x), \quad (5.2.119)$$

where the terms D and T are defined by

$$\begin{aligned} D(x)(\delta x) &= \{ \rho_u(\xi)(\delta y) + \rho_z(\xi)(\delta v) + \rho_q(\xi)(\delta p) \} \\ &+ \{ \rho_v(x)(\delta z) + \tilde{\rho}_y(x)(\delta u) + \tilde{\rho}_p(x)(\delta q) \} \end{aligned} \quad (5.2.120)$$

and

$$\begin{aligned} T(x)(\delta x) &= -\langle C''(u)(\delta u, v), R^{LS}(u) \rangle_Z - a''_{uu}(u, q)(\delta u, v, z) \\ &- a''_{qu}(u, q)(\delta u, p, z) - a''_{uq}(u, q)(\delta q, v, z) \end{aligned} \quad (5.2.121)$$

$$- a''_{qq}(u, q)(\delta q, p, z), \quad (5.2.122)$$

using the additional residuals:

$$\begin{aligned}\tilde{\rho}_y(x)(\psi) &:= \langle C'(u)(v), C'(u)(\psi) \rangle_Z - a'_u(u, q)(\psi, y), \\ \tilde{\rho}_p(x)(\sigma) &:= E'(q)(\sigma) - a'_q(u, q)(\sigma, y) + \alpha \langle p, \sigma \rangle_Q.\end{aligned}\tag{5.2.123}$$

The main idea is to replace the stationarity condition (5.2.102) by a simpler equation:

$$D(\tilde{x})(\delta x) = 0 \quad \forall \delta x \in X.\tag{5.2.124}$$

Similar to Proposition 5.2.2 we obtain:

PROPOSITION 5.2.4. *Let (u, q, z) be a stationary point of \mathcal{L} (2.3.37) and let $\tilde{H} = (\tilde{H}_{jk})$ be the reduced Gauß-Newton matrix defined by*

$$\tilde{H} = G^*G + \alpha I.\tag{5.2.125}$$

Further, let $\{w_j\}_{1 \leq j \leq n_p}$ be the tangent solutions defined in (2.3.30). Then, the solution $\tilde{x} = (u, q, z, \tilde{v}, \tilde{y}, \tilde{p}) \in X$ of (5.2.124) is given by:

$$\tilde{H}\tilde{p} = -\nabla E, \quad \tilde{v} = \sum_{j=1}^{n_p} w_j \tilde{p}_j,\tag{5.2.126}$$

and \tilde{y} is the solution of the following equation:

$$a'_u(u, q)(\phi, \tilde{y}) = \langle G\tilde{p}, C'(u)(\phi) \rangle_Z \quad \forall \phi \in V.\tag{5.2.127}$$

The proof of Proposition 5.2.4 is similar to the one of Proposition 5.2.2.

The computation of $(\tilde{v}, \tilde{y}, \tilde{p})$ is cheaper in comparison with (v, p, y) , because the matrix M is not required and the right hand side of (5.2.127) does not involve second derivatives of the semi-linear form a in contrast to (5.2.116). The corresponding error representation is formulated in the following theorem:

THEOREM 5.2.2. *Let (u, q, z) be the solution of the parameter identification problem and let $(\tilde{v}, \tilde{p}, \tilde{y})$ be defined by equations (5.2.126) and (5.2.127). Further we denote by*

$$\tilde{x}_h = (u_h, q_h, z_h, \tilde{v}_h, \tilde{p}_h, \tilde{y}_h)$$

the corresponding Galerkin solution. Then, the following error representation holds:

$$E(q) - E(q_h) = \frac{1}{2} \{ \rho_u(\xi_h)(\tilde{y} - i_h \tilde{y}) + \tilde{\rho}_y(\tilde{x}_h)(u - i_h u) \} + \tilde{R} + P,$$

where \tilde{R} is a remainder term defined by

$$\tilde{R} = \frac{1}{2} \int_0^1 D''(\tilde{x}_h + s\tilde{e})(\tilde{e}, \tilde{e}, \tilde{e})s(s-1)ds - \int_0^1 T'(\tilde{x}_h + s\tilde{e})(\tilde{e}, \tilde{e})sds,\tag{5.2.128}$$

with $\tilde{e} = \tilde{x} - \tilde{x}_h$. The additional remainder term P admits the estimate:

$$|P| \leq \tilde{C} (\|e_u\|_V + \|e_q\|_Q + \|\delta_h \tilde{v}\|_V + \|\delta_h \tilde{z}\|_V) \|R^{LS}(u)\|_Z, \quad (5.2.129)$$

where $e_u := u - u_h$, $e_q := q - q_h$, $\delta_h \phi := \phi - i_h \phi$ is an interpolation error operator on V and $R^{LS}(u)$ is the least squares residual defined in (2.1.3). The normalized adjoint solution $\tilde{z} \in V$ is determined by:

$$a'_u(u, q)(\phi, \tilde{z}) = \left\langle -\frac{R^{LS}(u)}{\|R^{LS}(u)\|}, C'(u)(\phi) \right\rangle_Z \quad \forall \phi \in V, \quad (5.2.130)$$

if the least squares residual $R^{LS}(u)$ does not vanish; otherwise we set $\tilde{z} = 0$. The constant \tilde{C} does not depend on the mesh parameter h nor on the measurements \tilde{C} .

PROOF. Similar to Proposition 5.2.1 we obtain the following error representation:

$$\begin{aligned} E(q) - E(q_h) &= \int_0^1 \mathcal{M}'(\tilde{x}_h + s\tilde{e})(\tilde{e}) ds \\ &= \int_0^1 D(\tilde{x}_h + s\tilde{e})(\tilde{e}) ds + \int_0^1 T(\tilde{x}_h + s\tilde{e})(\tilde{e}) ds. \end{aligned}$$

We approximate the first integral by the trapezoidal rule, the second by the box rule and obtain

$$E(q) - E(q_h) = \frac{1}{2}D(\tilde{x}_h)(\tilde{e}) + \frac{1}{2}D(\tilde{x})(\tilde{e}) + T(\tilde{x})(\tilde{e}) + \tilde{R}, \quad (5.2.131)$$

with the corresponding remainder term \tilde{R} defined in (5.2.128). The term $D(\tilde{x})(\tilde{e})$ vanishes, and due to Galerkin orthogonality the term $D(\tilde{x}_h)(\tilde{e})$ can be replaced by $D(\tilde{x}_h)(\tilde{x} - \hat{x}_h)$ with \hat{x}_h arbitrarily chosen. We set

$$\hat{x}_h = (i_h u, q, i_h z, i_h \tilde{v}, \tilde{p}, i_h \tilde{y}) \quad (5.2.132)$$

and obtain that the residuals ρ_q and $\tilde{\rho}_p$ vanish. We define the remainder term P by

$$P = T(\tilde{x})(\tilde{e}) + \frac{1}{2}\rho_z(\xi_h)(\delta_h z) + \frac{1}{2}\rho_v(\tilde{x}_h)(\delta_h \tilde{v}) \quad (5.2.133)$$

and it remains to prove the estimation (5.2.129). The term P has the following explicit form:

$$\begin{aligned} P &= -\langle C'''(u)(e_u, \tilde{v}), R^{LS}(u) \rangle_Z - a''_{uu}(u, q)(e_u, \tilde{v}, z) \\ &\quad - a''_{qu}(u, q)(e_u, \tilde{p}, z) - a''_{uq}(u, q)(e_q, \tilde{v}, z) \\ &\quad - a''_{qq}(u, q)(e_q, \tilde{p}, z) - \langle C'(u)(\delta_h \tilde{v}), R^{LS}(u) \rangle_Z \\ &\quad - a'_u(u, q)(\delta_h \tilde{v}, z) - a'_u(u, q)(\tilde{v}, \delta_h z) - a'_q(u, q)(\tilde{p}, \delta_h z). \end{aligned} \quad (5.2.134)$$

We use the fact that $\bar{z} = z \|R^{LS}(u)\|_Z$ and rewrite (5.2.134) as

$$\begin{aligned}
P &= -\langle C''(u)(e_u, \tilde{v}), R^{LS}(u) \rangle_Z - \|R^{LS}(u)\|_Z a''_{uu}(u, q)(e_u, \tilde{v}, \bar{z}) \\
&- \|R^{LS}(u)\|_Z a''_{qu}(u, q)(e_u, \tilde{p}, \bar{z}) - \|R^{LS}(u)\|_Z a''_{uq}(u, q)(e_q, \tilde{v}, \bar{z}) \\
&- \|R^{LS}(u)\|_Z a''_{qq}(u, q)(e_q, \tilde{p}, \bar{z}) + \langle C'(u)(\delta_h \tilde{v}), R^{LS}(u) \rangle_Z \\
&- \|R^{LS}(u)\|_Z a'_u(u, q)(\delta_h \tilde{v}, \bar{z}) - \|R^{LS}(u)\|_Z a'_u(u, q)(\tilde{v}, \delta_h \bar{z}) \\
&- \|R^{LS}(u)\|_Z a'_q(u, q)(\tilde{p}, \delta_h \bar{z}).
\end{aligned} \tag{5.2.135}$$

Using the Cauchy-Schwarz inequality and the continuity of the derivatives of C and a' completes the proof. \square

REMARK 5.2.5. For evaluation of the resulting error estimator, the local interpolation errors $u - i_h u$ and $\tilde{y} - i_h \tilde{y}$ are approximated by using higher order reconstructions. For the discussion of the localization of the error estimator, i.e. for the node-wise (cell-wise) representation of it, see Section 5.4.

REMARK 5.2.6. In the case of perfect match, i.e. $R^{LS}(u) = 0$, the remainder term P vanishes and the error estimators resulting from Theorem 5.2.1 and from Theorem 5.2.2 are identical.

5.3. An extension to more general error functionals

In this section we extend our a posteriori error estimation for more general error functionals $E : V \times Q \rightarrow \mathbb{R}$, i.e. for the case, if the quantity of interest depends on the state variable as well as on the parameters, see Vexler [87] and Becker & Vexler [23].

This extension allows a consideration of the following problem: We are given a mathematical model of a physical process, described by a system of partial differential equations. This model depends on a finite number of unknown (or just imprecisely known) model parameters, which cannot be measured directly. We are interested in the value of some quantity of the process (*quantity of interest* or *output*), which cannot be measured directly either (or whose measurement would require too much effort). This quantity of interest is given as a function of the state variable and parameters, denoted by E . As usual we are given a set of measurements and our aim is to compute the quantity of interest for the calibrated model with parameters, which fit the measurements.

One possible approach to this kind of problems is to split this problem in a pure parameter identification problem and an output-oriented simulation of the process. By this approach, first the unknown parameters are estimated from the given measurements using parameter identification techniques, and then the quantity of interest is computed by an appropriate simulation. But this approach has some drawbacks: First, the appropriate discretizations for the parameter identification problem and for the output-oriented simulation can be completely different; and second, the required accuracy of the parameters for achieving a given tolerance for the error in

the quantity of interest is a priori unknown due to the (in general unknown) derivatives of the quantity of interest with respect to the parameters.

We suggest another approach for more effective solution of this kind of problems. We consider this problem as a parameter identification problem and choose the finite element discretizations adaptively according to the quantity of interest. To this end, we derive an a posteriori error estimator, which aims at controlling the error in the quantity of interest. We use this error estimator within the adaptive mesh refinement algorithm described above, with purpose to improve the accuracy of the computed output in an efficient way.

Similar to the consideration in the previous section, there are two possibilities to derive the announced error estimator for the quantity of interest. In the following theorem we give the result, which requires the Hessian of the reduced cost functional as in Theorem 5.2.1.

THEOREM 5.3.1. *Let (u, q, z) be a solution of the parameter identification problem and (u_h, q_h, z_h) the corresponding Galerkin solutions. Let moreover H denote the reduced Hessian, see (2.3.34), and (v, p, y) be defined as follows:*

$$Hp = g, \quad (5.3.136)$$

where the components of g are:

$$\begin{aligned} g_j &= E'_{q_j}(u, q)(1) + E'_u(u, q)(w_j), \\ v &= \sum_{j=1}^{n_p} w_j p_j, \end{aligned} \quad (5.3.137)$$

and $y \in V$ is given by:

$$\begin{aligned} a'_u(u, q)(\phi, y) &= \langle Gp, C'(u)(\phi) \rangle_Z - \langle C''(u)(\phi, v), R^{LS}(u) \rangle_Z - a''_{uu}(u, q)(\phi, v, z) \\ &\quad - a''_{uq}(u, q)(\phi, p, z) + E'_u(u, q)(\phi) \quad \forall \phi \in V. \end{aligned} \quad (5.3.138)$$

Let the discrete variable (v_h, p_h, y_h) be defined in the similar way. Then, there holds the following error representation:

$$E(u, q) - E(u_h, q_h) = \frac{1}{2} \{ \rho_u(\xi_h)(y - i_h y) + \rho_z(\xi_h)(v - i_h v) \} \quad (5.3.139)$$

$$+ \frac{1}{2} \{ \rho_v(x_h)(z - i_h z) + \bar{\rho}_y(x_h)(u - i_h u) \} + R, \quad (5.3.140)$$

where ρ_u , ρ_z and ρ_v are defined as in (5.2.100) and in (5.2.112), and $\bar{\rho}_y$ is given by:

$$\begin{aligned} \rho_y(x)(\psi) &:= E'(u, q)(\psi) + \langle C'(u)(v), C'(u)(\psi) \rangle_Z - \langle C''(u)(\psi, v), R^{LS}(u) \rangle_Z \\ &\quad - a''_{uu}(u, q)(\psi, v, z) - a''_{uq}(u, q)(\psi, p, z) - a'_u(u, q)(\psi, y) \end{aligned} \quad (5.3.141)$$

and R is a cubic remainder term due to linearization.

PROOF. The proof is given using the same techniques as in Section 5.2. \square

REMARK 5.3.1. Similar to Theorem 5.2.2 one obtains the error estimator, which does not require the use of the Newton method.

5.4. Localization of the error estimator

For numerical evaluation of the error representation from Theorem 5.2.1, Theorem 5.2.2 and Theorem 5.3.1 we have to approximate the local interpolation errors, e.g. $u - i_h u$. In Becker & Rannacher [19], it is suggested to use a recovery process of the computed solution u_h by higher-order polynomials. In our practical realization, we exploit the patch-wise structure of finite element meshes, see Section 3.1, and use interpolation in the space of higher order finite elements $V_{2h}^{(2)}$ on patches \mathcal{T}_{2h} .

REMARK 5.4.1. In the following, we concentrate on the consideration of bilinear finite elements V_h and biquadratic elements $V_{2h}^{(2)}$ on patches such that there holds:

$$\dim V_{2h}^{(2)} = \dim V_h.$$

Let $i_{2h}^{(2)} : V_h \rightarrow V_{2h}^{(2)}$ be the corresponding interpolation operator; the interpolation errors are numerically approximated by:

$$u - i_h u \approx i_{2h}^{(2)} u_h - u_h.$$

Moreover, for using the resulting error estimator in the adaptive mesh refinement algorithm, we have to localize the estimator to the cell-wise or node-wise contributions. In the numerical examples described below we use node-wise localization. A direct localization of the terms from the error representations, such

$$\rho_y(x_h)(i_{2h}^{(2)} u_h - u_h),$$

leads, in general, to the local contributions of wrong order (overestimation) due to oscillatory behavior of the residual terms. To overcome this, we introduce a so called filtering operator $\pi_h : V_h \rightarrow V_h$ by setting:

$$\pi_h \xi := \xi - i_{2h}^h \xi,$$

where by i_{2h}^h we denote the nodal interpolation operator $i_{2h}^h : V_h \rightarrow V_{2h}$. The space V_{2h} is the bilinear finite element space on patches. This construction exploits the patch-structure of the mesh, which implies, that the finite element spaces V_{2h} and V_h are nested, i.e. $V_{2h} \subset V_h$. Due to the fact, that $V_{2h} \subset V_{2h}^{(2)}$, we obtain using Galerkin orthogonality:

$$\rho_y(x_h)(i_{2h}^{(2)} u_h - u_h) = \rho_y(x_h)(i_{2h}^{(2)} \pi_h u_h).$$

The last term is used in the following localization procedure: We denote the nodal basis of V_h by:

$$\{\phi_i\}, \quad i = 1 \dots n,$$

where $n = \dim V_h$. Further, for each node we introduce a quadratic nodal function $\phi_i^{(2)}$, which is given by:

$$\phi_i^{(2)} = i_{2h}^{(2)} \phi_i \in V_{2h}^{(2)}.$$

Let now $\Phi_i^y \in \mathbb{R}$ denote the contribution of the residual $\rho_y(x_h)$ with respect to the quadratic function $\phi_i^{(2)}$, i.e.

$$\Phi_i^y = \rho_y(x_h)(\phi_i^{(2)}).$$

For the filtered primal solution $\pi_h u_h$ the nodal values are denoted by U_i^π such that there holds:

$$\pi_h u_h = \sum_{i=1}^n U_i^\pi \phi_i.$$

This allows the following representation:

$$\rho_y(x_h)(i_{2h}^{(2)} \pi_h u_h) = \sum_{i=1}^n U_i^\pi \Phi_i^y.$$

According to this, the computable local quantities for each node are :

$$U_i^\pi \Phi_i^y, \quad i = 1 \dots n.$$

The contributions of other residual are computed in the same way. The error representation from Theorem 5.2.2 lead to the following local quantities:

$$\eta_{h,i} = \frac{1}{2} |U_i^\pi \Phi_i^y + Y_i^\pi \Phi_i^u|,$$

where Y_i^π and Φ_i^u are defined analog by

$$\Phi_i^u = \rho_u(x_h)(\phi_i^{(2)})$$

and

$$\pi_h y_h = \sum_{i=1}^n Y_i^\pi \phi_i.$$

The value of the error estimator is computed by:

$$\eta_h = \frac{1}{2} \left| \sum_{i=1}^n (U_i^\pi \Phi_i^y + Y_i^\pi \Phi_i^u) \right|$$

and there holds:

$$\eta_h \leq \sum_{i=1}^n \eta_{h,i}.$$

For more details of this procedure see Braack [28] and Braack & Ern [29]. For using the node-wise contributions $\eta_{h,i}$ in a refinement procedure, the information has to be shifted to cells. Thereafter, standard algorithms can be used, see [18, 19].

For discussion of other localization techniques we refer to Becker & Rannacher [18, 19]. Therein, the residual terms are cell-wise integrated by parts resulting in cell-wise contributions.

5.5. Numerical examples

In this section, we illustrate the usage of the a posteriori error estimators described above for some prototypical two-dimensional elliptic problems. Throughout, the discretization of the state equation uses piecewise bilinear finite elements on meshes consisting of quadrilaterals. The resulting nonlinear state equations are solved by Newton method and the solution of the linear subproblems are computed using a multigrid algorithm on locally refined meshes, see Becker & Braack [13]. With these ingredients, the total numerical cost for solution on a given mesh behaves like $O(N)$, where N is the number of nodes. All computations are done on the basis of the package *ParamGascoigne* for treating parameter identification problems. This C++ library is implemented by the author as an extension of the finite element toolkit *Gascoigne3D*, see [12].

5.5.1. EXAMPLE 1

We consider a convection-diffusion equation with unknown constant transport direction (q_1, q_2) in the unit square $\Omega = (0, 1)^2$ (as in the example in Section 4.4):

$$\begin{aligned} -\Delta u + q_1 u_x + q_2 u_y &= 2 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{5.5.142}$$

The parameters $(q_1, q_2) = (8, 8)$ are estimated using measurements given by the values of the state variable at five different points ξ_i as in the example in Section 4.4, see (4.4.84). The components of the corresponding observation operator C have the following form:

$$C_i(v) = v(\xi_i), \tag{5.5.143}$$

and the parameter identification problem is formulated as follows: For $(u, q) \in V \times Q$ with $V = H_0^1(\Omega)$ and $Q = \mathbb{R}^2$

$$\text{Minimize } \frac{1}{2} \sum_{i=1}^5 (u(\xi_i) - \bar{C}_i)^2 \tag{5.5.144}$$

under the constraint (5.5.142), where \bar{C}_i denote the components of the measurement vector $\bar{C} \in Z = \mathbb{R}^5$ and are given by the values of the state variable u for the exact parameter q , i.e. $\bar{C}_i = u(\xi_i)$.

Estimating the error in parameter

We estimate the error in the first component of $q \in Q$, i.e. we set:

$$E_1(q) := q_1. \quad (5.5.145)$$

First, we study the quality of the generated locally refined meshes. For comparison we also consider global refinement and refinement by an “energy” estimator for the state variable, which aims to control the error $u - u_h$ in $H^1(\Omega)$. Figure 5.5.1 compares the accuracy achieved on meshes resulting from these three types of refinement.

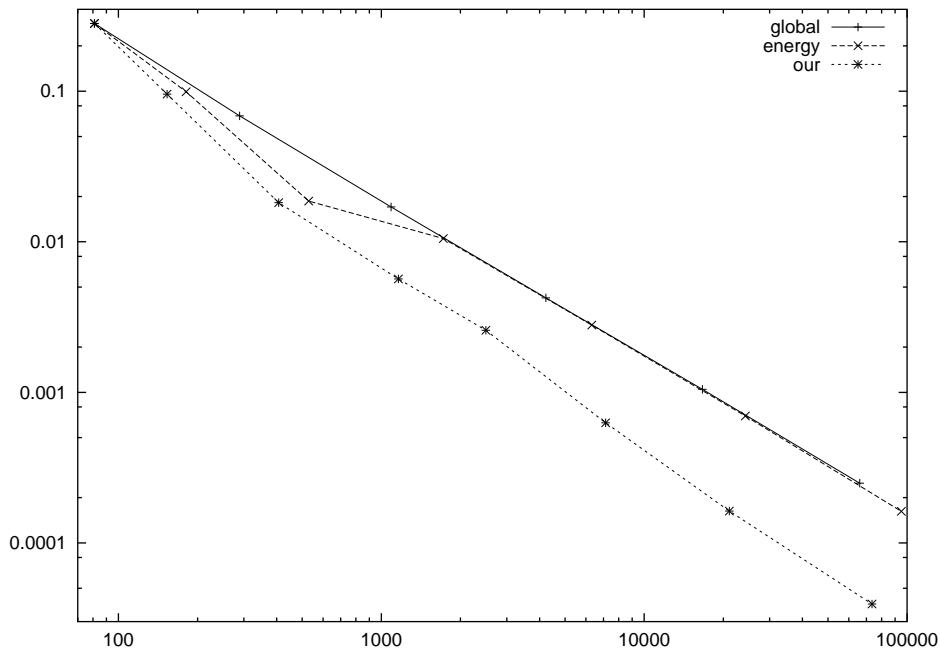


FIGURE 5.5.1. Errors in q_1 for different refinement strategies vs. number of nodes.

As seen from Figure 5.5.1, adaptive refinement based on the “energy”-estimator leads to almost the same reduction of the error as global refinement. However, the strategy based on our estimator leads to an obvious saving in the number of unknowns necessary to achieve a prescribed accuracy level.

Next, we investigate the quantitative behavior of the estimator which is important for error control. Comparison between the error in q_1 and the value of the estimator is shown in

Table 5.5.1. There, the effectivity index of the error estimator is defined by

$$I_{eff} := (E(q) - E(q_h))/\eta_h. \quad (5.5.146)$$

TABLE 5.5.1. Efficiency of the error estimator for the error functional $E_1(q) = q_1$.

N	$E_1(q) - E_1(q_h)$	η_h	I_{eff}
81	2.80e-1	3.60e-1	0.78
153	9.59e-2	1.04e-1	0.92
407	1.82e-2	1.99e-2	0.91
1161	5.66e-3	5.39e-3	1.05
2487	2.47e-3	2.41e-3	1.02
7129	6.27e-4	6.14e-4	1.02
21091	1.63e-4	1.60e-4	1.02

Computations for Table 5.5.1 are done on the same sequence of locally refined meshes as before, see Figure 5.5.1.

Some typical meshes resulting from application of the a posteriori error estimator for the error functionals $E_1(q) = q_1$, $E_2(q) = q_2$, and $E_3(q) = q_1^2 + q_2^2$ are shown in Figure 5.5.2.

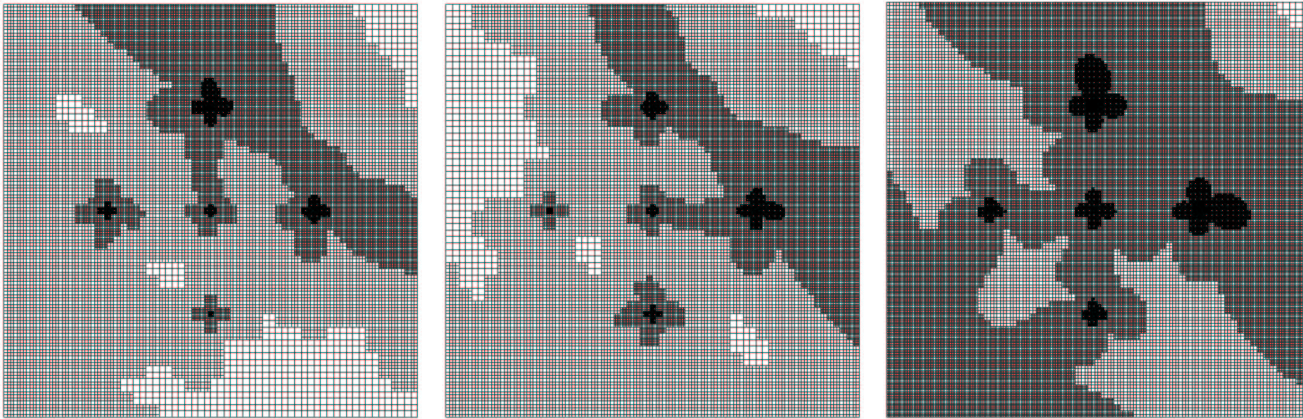


FIGURE 5.5.2. Typical meshes produced for $E_1(q) = q_1$ (left), $E_2(q) = q_2$ (middle) and $E_3(q) = q_1^2 + q_2^2$ (right).

REMARK 5.5.1. The presented example with the error functional $E_1(q) = q_1$ mimics a typical situation. One distinguishes between a set of parameters which are of primary interest, q_1 , and others, which are also unknown and determined on the way. Depending on the choice of primary parameters, the meshes may differ significantly.

Estimating the error in the quantity of interest

For illustrating the error estimation, discussed in Section 5.3, we consider the quantity of interest given by the mean value

$$E_{\Omega_0}(u) = \int_{\Omega_0} u \, dx \quad (5.5.147)$$

on a subdomain Ω_0 defined by

$$\Omega_0 = \left(\frac{15}{16}, 1\right) \times \left(0, \frac{1}{16}\right).$$

We test again the efficiency of the error estimator and the quality of the generated meshes. In Table 5.5.2 the error, the estimation of it and the effectivity index of the error estimator, defined as in (5.5.146) are shown.

TABLE 5.5.2. Efficiency of the error estimator for the error functional E_{Ω_0} .

N	$E_{\Omega_0}(u) - E_{\Omega_0}(u_h)$	η_h	I_{eff}
81	1.46e-0	4.07e-0	0.36
163	1.65e-1	8.10e-2	2.04
375	4.13e-2	4.24e-2	0.97
1009	1.07e-2	1.07e-2	1.00
2463	2.55e-3	2.56e-3	0.99
6135	7.65e-4	7.79e-4	0.98
16713	1.92e-4	1.95e-4	0.98

In Figure 5.5.3 we compare the accuracy with respect to the quantity of interest achieved on the meshes resulting from the application of the error estimator for E_{Ω_0} with some other types of mesh refinement. To this end, we consider uniform mesh refinement, refinement corresponding to the error estimator derived for the error in parameters ($E_s(q) = q_1 + q_2$) and for the error in the cost functional ($J(u) = \frac{1}{2}\|C(u) - \bar{C}\|^2$). The latter corresponds to the a posteriori error estimator for optimization problems, see Becker & Rannacher [19] and Becker, Kapp & Rannacher [20].

As seen from Figure 5.5.3 the strategy of the a posteriori error estimation derived above is the most efficient one. The strategies, which does not take into account the quantity of interest, i.e. the error functional E_{Ω_0} , are even worse than the uniform mesh refinement. In Figure 5.5.4 we show typical meshes resulting from the a posteriori error estimator for E_{Ω_0} and from the estimator for the cost functional $J(u)$.

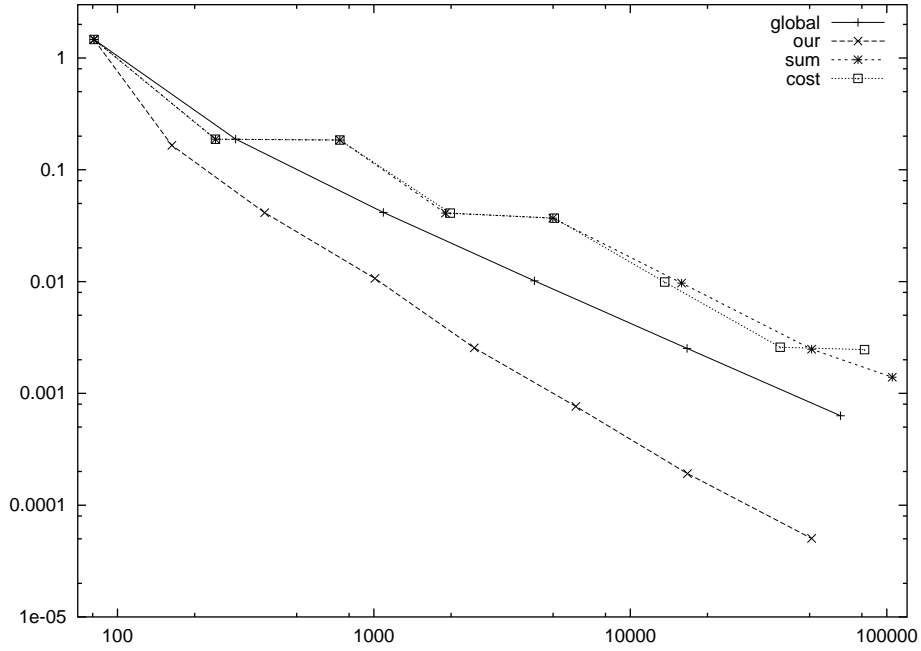


FIGURE 5.5.3. Errors in E_{Ω_0} for different refinement strategies vs. number of nodes.

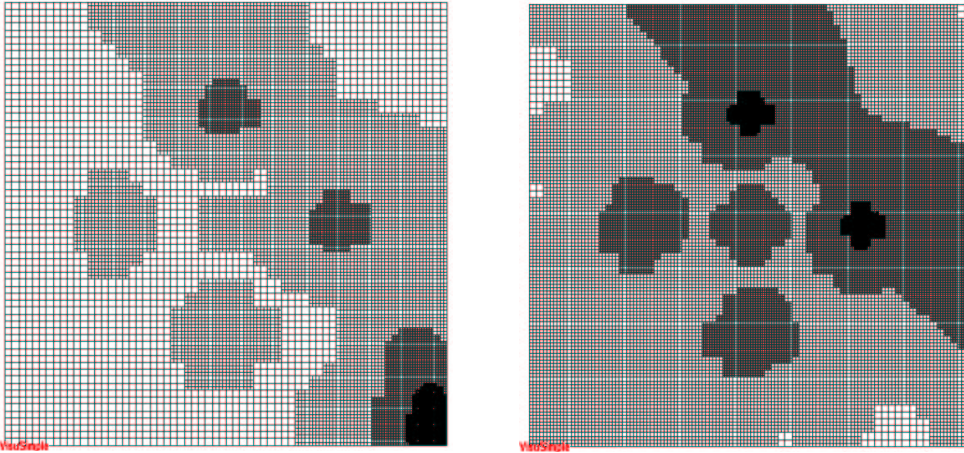


FIGURE 5.5.4. Typical meshes produced by a posteriori error estimator for E_{Ω_0} (left) and for the cost functional $J(u)$ (right).

5.5.2. EXAMPLE 2

In this example we investigate a simple case of non-vanishing least squares residual. Here, we consider a convection-diffusion-reaction equation with a nonlinear reaction term:

$$\begin{aligned} -\Delta u + e^{q_1 u} + q_2 u_x + q_3 u_y &= 2 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega. \end{aligned} \quad (5.5.148)$$

Again, we choose $\Omega = (0, 1)^2$. The parameters $(q_1, q_2, q_3) \in Q = \mathbb{R}^3$ are estimated using measurements given by the values of the state variable at the same points ξ_i (4.4.84) as in Example 1. The parameter identification problem is formulated as follows: For $(u, q) \in V \times Q$ with $V = H_0^1(\Omega)$ and $Q = \mathbb{R}^3$

$$\text{Minimize } \frac{1}{2} \sum_{i=1}^5 (u(\xi_i) - \bar{C}_i)^2 + \frac{\alpha}{2} \|q - \bar{q}\|_Q^2 \quad (5.5.149)$$

under the constraint (5.5.148), where \bar{C}_i denote the components of the measurement vector $\bar{C} \in Z = \mathbb{R}^5$ and $\bar{q} \in Q$ is a given reference parameter. In contrast to Example 1 we consider the case of $\alpha > 0$. The data $(\bar{C}, \bar{q}$ and $\alpha)$ are chosen, such that the least squares residual $R^{LS}(u)$ does not vanish.

Here, the a posteriori error estimators η_1 and η_2 based on the error representation from Theorem 5.2.1 and Theorem 5.2.2 respectively, differ. In order to illustrate the difference between the estimators, we consider two cases with different magnitude of $\|R^{LS}(u)\|_Z$. In both cases the error functional is $E(q) = q_2$. The values of the estimators η_1 and η_2 are shown in Table 5.5.4 and Table 5.5.3. The corresponding effectivity indexes I_{eff}^1 and I_{eff}^2 are defined as in (5.5.146).

TABLE 5.5.3. Comparison between the error in $E(q)$ and the values of the estimators η_1 and η_2 for the data with $\|R^{LS}(u)\|_Z \approx 3 \cdot 10^{-4}$

N	$E(q) - E(q_h)$	η_1	I_{eff}^1	η_2	I_{eff}^2
25	3.27e-1	6.92e-2	4.72	4.87e-2	6.73
81	7.53e-2	8.53e-2	0.88	7.80e-2	0.96
289	1.84e-2	1.87e-2	0.98	1.74e-2	1.06
1089	4.57e-3	4.51e-3	1.01	4.21e-3	1.09
4225	1.14e-3	1.12e-3	1.02	1.04e-3	1.10

TABLE 5.5.4. Comparison between the error in $E(q)$ and the values of the estimators η_1 and η_2 for the data with $\|R^{LS}(u)\|_Z \approx 4 \cdot 10^{-2}$

N	$E(q) - E(q_h)$	η_1	I_{eff}^1	η_2	I_{eff}^2
25	1.87e-1	2.06e-1	0.90	5.31e-2	3.52
81	3.58e-2	4.74e-2	0.76	1.26e-2	2.84
289	8.24e-3	8.71e-3	0.94	3.29e-3	2.50
1089	1.99e-3	2.00e-3	0.99	8.31e-4	2.39
4225	4.70e-4	4.85e-3	0.97	2.08e-4	2.26

5.5.3. EXAMPLE 3

In this section we compare our general approach to mesh refinement for parameter identification problems with some heuristic mesh refinement algorithms.

We consider three types of heuristic approaches to mesh refinement: a strategy based only on the information obtained from the computed state variable, a strategy based only on the a priori knowledge of the structure of the observation operator and a strategy, which combines both types of information.

For this comparison, we consider the following diffusion-reaction equation with unknown coefficient q in the unit square $\Omega = (0, 1)^2$:

$$\begin{aligned} -q\Delta u + su &= 2 && \text{in } \Omega, \\ u &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{5.5.150}$$

where s is chosen as $s = 200$.

The parameter q is estimated using measurements given by the values of the state variable at nine different points ξ_i , see Figure 5.5.5. The exact value of the parameter is $q = 1$.

REMARK 5.5.2. We note, that the chosen values of s and q lead to boundary layer behavior of the solution of the state equation (5.5.150).

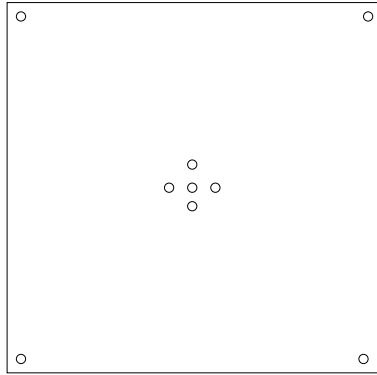


FIGURE 5.5.5. The computational domain with measurement points marked by circles.

The components of the corresponding observation operator $C(\cdot)$ have the following form:

$$C_i(v) = v(\xi_i), \quad i = 1, 2, \dots, 9.$$

and the parameter identification problem is formulated as follows: For $(u, q) \in V \times Q$ with $V = H_0^1(\Omega)$ and $Q = \mathbb{R}$

$$\text{Minimize } \frac{1}{2} \sum_{i=1}^9 (u(\xi_i) - \bar{C}_i)^2 \tag{5.5.151}$$

under the constraint (5.5.150). The components \bar{C}_i of the measurement vector $\bar{C} \in Z = \mathbb{R}^9$ are given by the values of the state variable u for the exact parameter q , i.e.

$$\bar{C}_i = u(\xi_i), \quad i = 1, 2 \dots 9.$$

First, we compare the quality of meshes generated by our a posteriori error estimator for this problem with a typical strategy based on a posteriori information obtained by the state variable, i.e with the mesh refinement guided by one of the well-known “energy” type error estimators for uncontrolled equation, see e.g. Bank & Weiser [8] and Babuška & Miller [6]. This estimator aims to control the error $u - u_h$ in H^1 -norm, but it does not take care of the structure of the parameter identification problem.

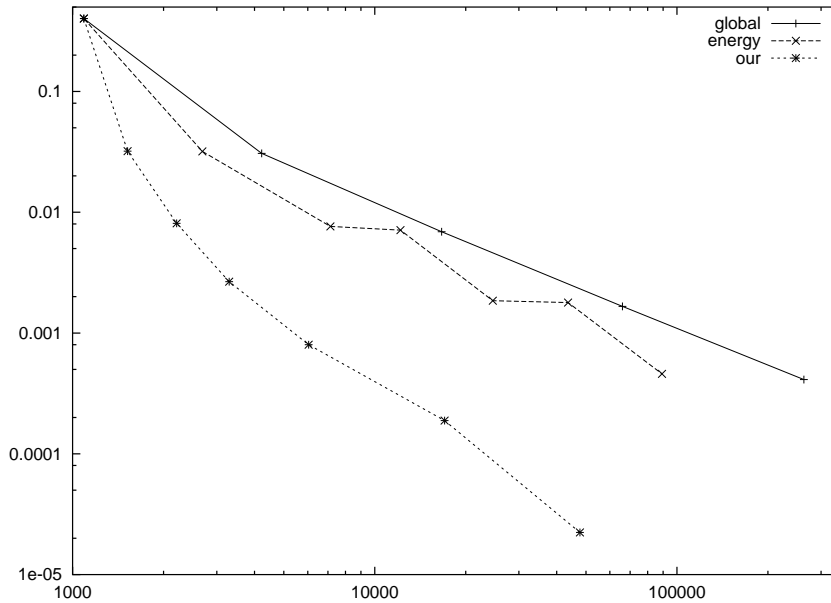


FIGURE 5.5.6. Errors in q for different refinement strategies vs. number of nodes (global refinement, “energy”-based refinement and refinement resulting from our a posteriori error estimator).

As seen from Figure 5.5.6, adaptive refinement based on the “energy” estimator leads to the similar reduction of the error as global refinement. However, the strategy based on our error estimator leads to an obvious saving in the number of unknowns necessary to achieve a prescribed accuracy level.

Next, we compare our strategy for mesh refinement with the following heuristic approach: In each iteration of the mesh refinement we refine the cells, which lie close to one of the

measurement points, i.e. in the set

$$\bigcup_i \{x \mid \|x - \xi_i\| \leq r\}, \quad (5.5.152)$$

where $r \in \mathbb{R}_+$ is a given number. In contrast to our approach, this strategy is unable to weight the relative importance of the measurement points. The corresponding comparison with our strategy is done in Figure 5.5.7 for two choices of r ($r = 0.04$ and $r = 0.1$).

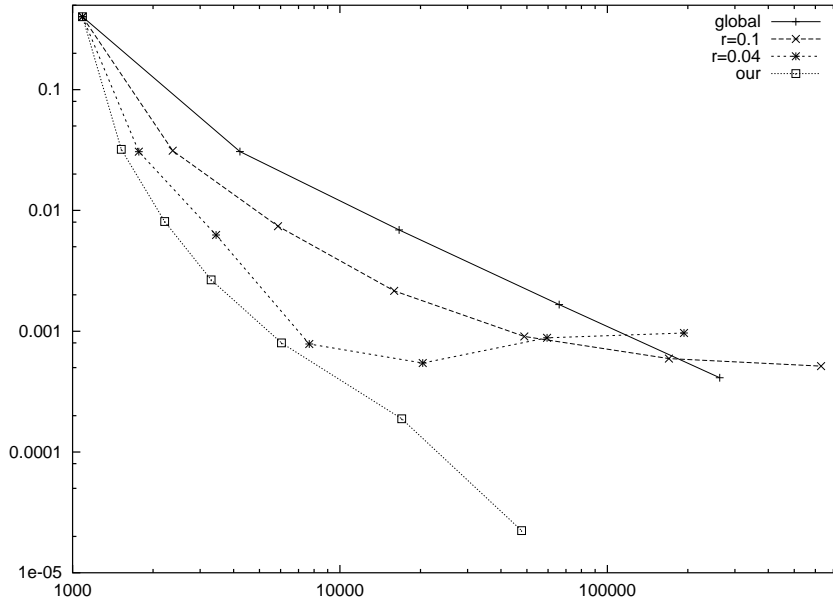


FIGURE 5.5.7. Errors in q for different refinement strategies vs. number of nodes (global refinement, refinement across measurement points and refinement resulting from our a posteriori error estimator).

After several steps one does not observe any error reduction despite increasing the number of nodes for both choices of r in the described strategy, as could be expected. Typical meshes resulting from application of our a posteriori error estimator, “energy” estimator and the last strategy are shown in Figure 5.5.8.

We also compare our mesh refinement procedure with a combination of both last heuristic methods. By this strategy both cells marked by “energy” estimator and cells across the measurement points (5.5.152) are refined. The corresponding comparison with our mesh refinement procedure is made in Figure 5.5.9 for two choices of r ($r = 0.04$ and $r = 0.1$).

The typical meshes resulting from application of this strategy for two choices of r ($r = 0.04$ and $r = 0.1$) are shown in Figure 5.5.10.

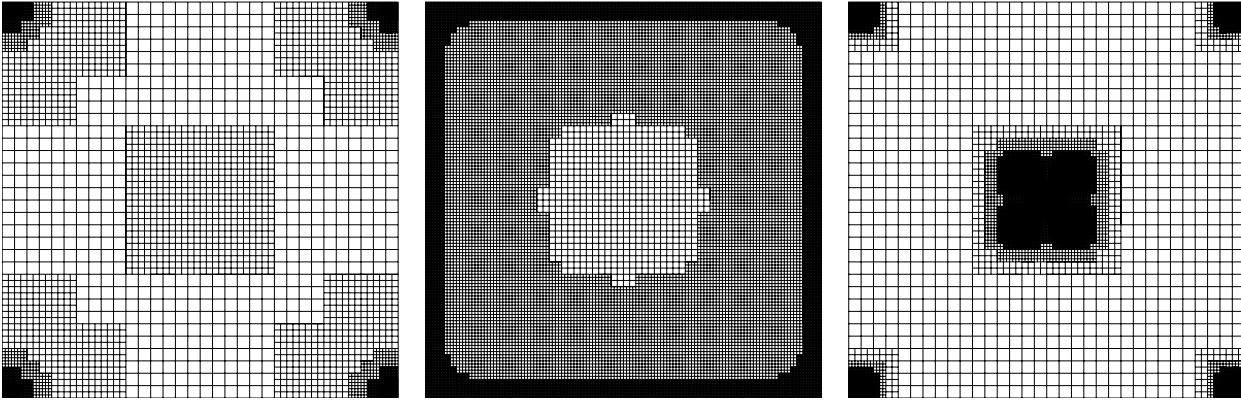


FIGURE 5.5.8. Typical meshes produced by our a posteriori error estimator (left), energy error estimator (middle) and the refinement across the measurement points (right).

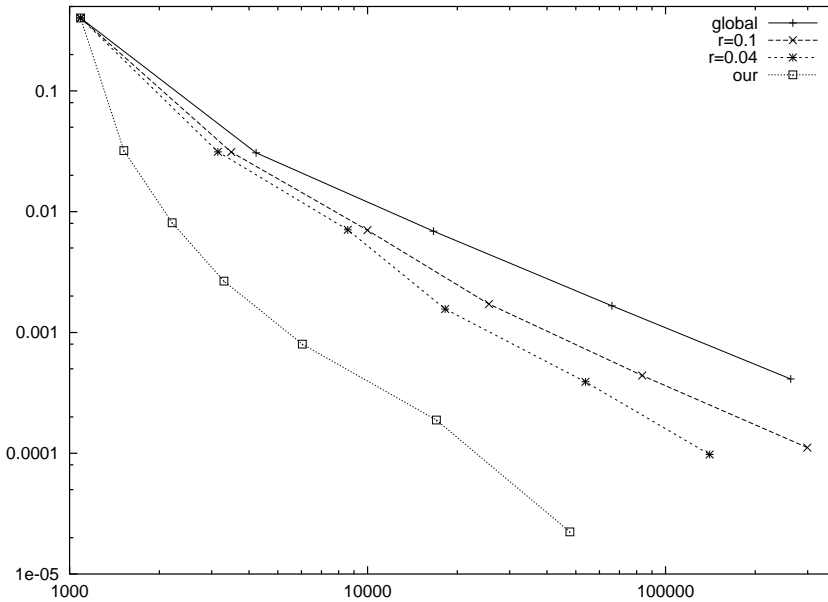


FIGURE 5.5.9. Errors in q for different refinement strategies vs. number of nodes (global refinement, refinement produced by combining the refinement across measurement points and “energy”-based refinement, and refinement resulting from our a posteriori error estimator).

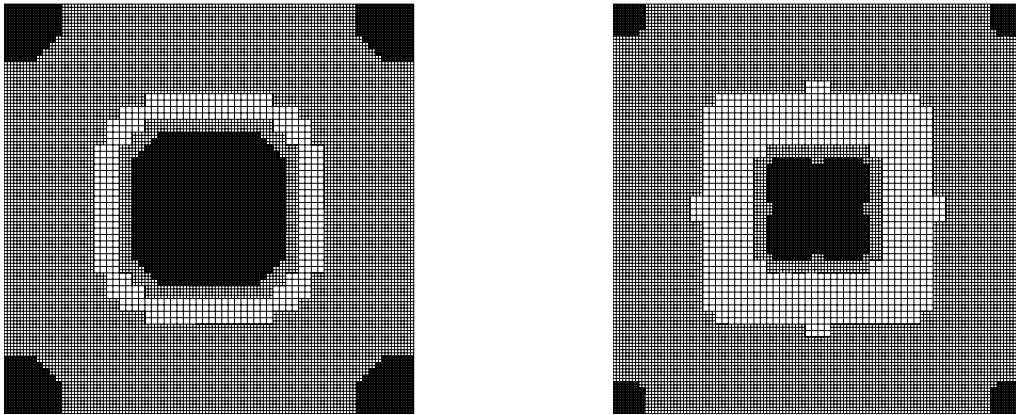


FIGURE 5.5.10. Typical meshes produced by combining the refinement according to “energy” error estimator and the refinement across the measurement points for $r = 0.1$ (left) and $r = 0.04$ (right).

CHAPTER 6

Application to CFD Problems

In this chapter we illustrate the capability of our approach on two flow problems, where the state is described by the incompressible Navier-Stokes equations. In Section 6.1 we describe the discretization of the Navier-Stokes equations by stabilized finite elements of equal order. The approach to stabilization is based on local projection techniques. Thereafter, in Section 6.2 an extension of the provided parameter identification techniques to the parameter-dependent boundary conditions is discussed. The application of this approach to covering unknown boundary conditions by CFD problems is discussed on two example problems. In Section 6.3 we consider a flow in a system of tubes describing a simple model for blood flow in arteries (bypass simulation). The unknown boundary conditions of Neumann type are parameterized, and the parameters are estimated using measurements of fluxes. A problem with parameterized Dirichlet boundary condition is discussed in Section 6.4. Here, we consider a flow around a cylinder in a channel with rough wall. The computation is done on a channel with even walls, and the roughness is modeled by prescription of parameterized inhomogeneous Dirichlet boundary conditions.

6.1. Discretization of the Navier-Stokes equations

This section is devoted to the discretization of the (stationary) Navier-Stokes equations with finite element method. Here, we use the discretization and stabilization techniques similar to the approach described in Becker [11] and Becker & Braack [14].

Let $\Omega \subset \mathbb{R}^2$ be a Lipschitz domain. Its boundary $\partial\Omega$ is divided into a part Γ_D where Dirichlet boundary conditions are subscribed and a part Γ_N , where Neumann-type conditions are given, $\partial\Omega = \Gamma_D \cup \Gamma_N$. The equations for the unknown velocity field v and pressure p read: Find $u = (v, p)$ such that

$$\begin{aligned} -\nu\Delta v + v \cdot \nabla v + \nabla p &= f && \text{in } \Omega, \\ \nabla \cdot v &= 0 && \text{in } \Omega, \\ v &= g && \text{on } \Gamma_D, \\ \nu \frac{\partial v}{\partial n} - pn &= p_0 n && \text{on } \Gamma_N, \end{aligned} \tag{6.1.153}$$

where $g \in H^{\frac{1}{2}}(\Gamma_D)$ describes the Dirichlet, $p_0 \in L^2(\Gamma_N)$ the Neumann boundary condition data, and n denote the outward unit normal vector to Γ_N . For the analysis of this type of Neumann

boundary conditions, see Heywood, Rannacher & Turek [63].

The starting point for any finite element discretization of the Navier-Stokes equations (6.1.153) is a variational formulation. The space of test function for the velocity is

$$H = \{\psi \in H^1(\Omega)^2 \mid \psi = 0 \text{ on } \Gamma_D\}.$$

Let $\hat{v} \in H^1(\Omega)^2$ be a function representing the Dirichlet boundary data, i.e.

$$\hat{v} = g \quad \text{on } \Gamma_D.$$

Further, we set $\hat{u} = (\hat{v}, 0)$, $V = H \times L^2(\Omega)$ and $\hat{V} = \hat{u} + V$. The semilinear form $a : \hat{V} \times V \rightarrow \mathbb{R}$ is defined by:

$$a(u)(\phi) = \nu(\nabla v, \nabla \psi) + (v \cdot \nabla v, \psi) - (p, \nabla \cdot \psi) + (\nabla \cdot v, \xi), \quad (6.1.154)$$

where $\phi = (\psi, \xi) \in V$ denotes the test functions for velocity and pressure. The corresponding weak formulation reads: Find $u \in \hat{V}$ such that

$$a(u)(\phi) = (f, \psi) + (p_0 n, \psi)_{\Gamma_N} \quad \forall \phi \in V. \quad (6.1.155)$$

The problem (6.1.155) is discretized using conforming finite elements on shape-regular quadrilateral meshes described in Section 3.1. We use finite elements of equal order for each component constructed in the similar way as in Section 3.1, i.e.:

$$V_h = \{\phi_h = (\psi_h, \xi_h) \in V \mid \phi_{h|K} \in \tilde{Q}_1(K)^3, K \in \mathcal{T}_h\},$$

where $\tilde{Q}_1(K)$ consists of shape functions obtained via a bilinear transformation from the space of bilinear functions $Q_1(\hat{K})$ on the reference cell $\hat{K} = (0, 1)^2$. Using the patch-structure of the mesh, the finite element space V_{2h} on the mesh \mathcal{T}_{2h} is defined in a similar way, cf. Figure 3.1.2 in Section 3.1.

It is well known, see e.g. Girault & Raviar [54], that the use of equal-order finite elements leads to an unstable Galerkin formulation, since the stiff pressure-velocity coupling for incompressible flows enforces spurious pressure modes. Moreover, the standard Galerkin discretization of the (nonlinear) convection term leads to oscillations, see e.g. Johnson [68]. Therefore, we add stabilization terms to the semilinear form a (6.1.154) in order to obtain a stable formulation.

Due to the patch-structure of the mesh, see Section 3.1, the finite element spaces V_{2h} and V_h are nested, i.e. $V_{2h} \subset V_h$. By i_{2h}^h we denote the nodal interpolation operator $i_{2h}^h : V_h \rightarrow V_{2h}$. By

$$\pi_h : V_h \rightarrow V_h, \quad \pi_h \xi = \xi - i_{2h}^h \xi$$

we denote the difference between the identity and this interpolation. For a velocity field β and a discrete state $u_h = (v_h, p_h) \in V_h$ we define a stabilization bilinear form

$$s(\beta)(u_h, \phi_h) = \sum_{K \in \mathcal{T}_h} \{ \alpha_K (\nabla(\pi_h p_h), \nabla(\pi_h \xi_h))_K + \delta_K (\beta \cdot \nabla(\pi_h v_h), \beta \cdot \nabla(\pi_h \psi_h))_K \},$$

where the cell-wise coefficients δ_K and α_K are chosen:

$$\delta_K := \frac{\delta_0 h_K^2}{6\nu + h_K |\beta|_K}, \quad \alpha_K = \alpha_0 h_K^2 / \nu.$$

Here, h_K denotes the diameter of the cell K and $|\beta|_K$ is the cell-wise value of the velocity field β . The parameters δ_0 and α_0 are usually chosen between 0.2 and 1. Let \hat{u}_h be a discrete approximation of the boundary data \hat{u} , then the stabilized discrete formulation reads: Find $u_h = (v_h, p_h) \in \hat{u}_h + V_h$ such that:

$$a(u_h)(\phi_h) + s(v_h)(u_h, \phi_h) = (f, \psi_h) + (p_0 n, \psi_h)_{\Gamma_N} \quad \forall \phi_h \in V_h.$$

For the analysis of stability and convergence of the described approximation scheme see Becker & Braack [14]. Similar stabilization techniques for the convective term are analyzed in Guermond [59].

6.2. Treatment of parameter-dependent boundary conditions

In this section we discuss an extension of the methods described in previous chapters to covering unknown boundary conditions. The modeling of boundary conditions is a typical difficulty for a broad class of CFD problems. We consider the case, where boundary conditions g and p_0 (6.1.153) depend on the unknown parameters $q \in Q$, i.e.

$$g = g(q), \quad p_0 = p_0(q).$$

Such a parameterization allows for formulating a problem with unknown boundary conditions as an parameter identification problem. Using the information obtained by measurements the unknown parameters can be estimated by application of the methods described in previous chapters. This allows for finding boundary conditions, which have the best possible match with the experiments.

If the parameters enter the Neumann type boundary condition, the computation of the derivatives with respect to parameters is done in the same way, as described in Propositions 2.3.1 and 2.3.2. This is due to the fact, that in this case the parameters are directly involved in the semilinear form a . The case of parameter-dependent Dirichlet boundary conditions requires some additional remarks. Suppose the state equation with inhomogeneous parameter-dependent Dirichlet boundary conditions is given by: Find $u \in \hat{u}(q) + V$ such that

$$a(u, q)(\phi) = (f, \phi) \quad \forall \phi \in V. \tag{6.2.156}$$

Let S be the solution operator of (6.2.156), see Proposition 2.2.1, and the reduced observation operator c be defined as in (2.3.23) by:

$$c(q) = C(S(q)).$$

Then there holds the following proposition describing the derivatives of the reduced observation operator c with respect to parameters:

PROPOSITION 6.2.1. *There holds:*

$$\frac{\partial c_i}{\partial q_j}(q) = G_{ij} = C'_i(u)(w_j), \quad i = 1 \dots n_m, \quad j = 1 \dots n_p,$$

with $u = S(q)$, C_i and c_i denote the components of the observation and the reduced observation operators respectively. The tangent solution w_j is sought in $\hat{u}'_{q_j}(q)(1) + V$ and is determined by:

$$a'_u(u, q)(w_j, \phi) = -a'_{q_j}(u, q)(1, \phi) \quad \forall \phi \in V.$$

PROOF. The proof is given similar to the one of Proposition 2.3.1. \square

6.3. Bypass simulation with unknown boundary conditions

In this section we consider a simplified model for the blood flow in arteries (bypass simulation). Here, a typical difficulty is the prescription of in- and outflow boundary conditions on $\Gamma_{in}^{(1)}$, $\Gamma_{in}^{(2)}$ and Γ_{out} respectively, see the configuration in Figure 6.3.1.

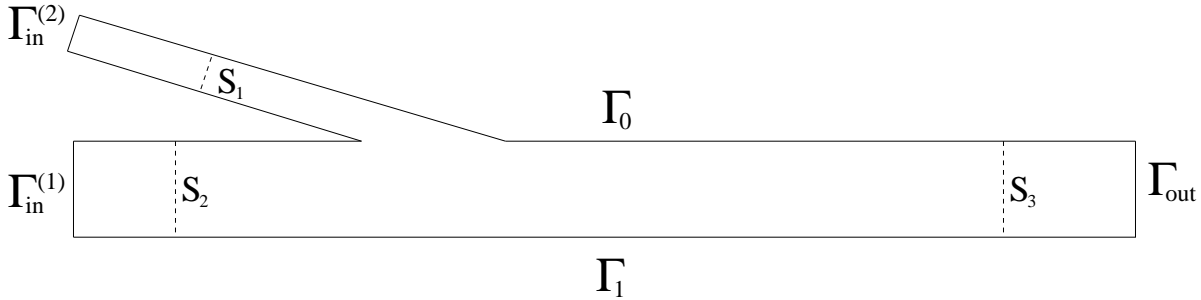


FIGURE 6.3.1. The configuration for the bypass simulation.

We parameterize these unknown boundary conditions by the parameter $q \in Q = \mathbb{R}^2$ in the following way:

$$\begin{aligned} \nu \frac{\partial v}{\partial n} - pn &= q_1 n && \text{on } \Gamma_{in}^{(1)}, \\ \nu \frac{\partial v}{\partial n} - pn &= q_2 n && \text{on } \Gamma_{in}^{(2)}, \\ \nu \frac{\partial v}{\partial n} - pn &= 0 && \text{on } \Gamma_{out}, \\ v &= 0 && \text{on } \Gamma_0 \cup \Gamma_1. \end{aligned}$$

This parameterization can be interpreted as follows: The parameters q_1 and q_2 describe the pressure difference between $\Gamma_{in}^{(1)}$ and Γ_{out} , and between $\Gamma_{in}^{(2)}$ and Γ_{out} respectively, cf. Heywood, Rannacher & Turek [63].

We assume, the measurements to be given by fluxes at S_1 , S_2 and S_3 , see Figure 6.3.1. The corresponding observation operator is given by:

$$C_i(u) = \int_{S_i} v \cdot n \, ds, \quad i = 1, 2, 3.$$

The choice of the type of measurements is not crucial for this application. The measurements of point values of the velocity or of the pressure are also possible.

The unknown parameter are estimated from a set of measurements $\bar{C} \in Z = \mathbb{R}^3$. The solution of the state equation for the exact parameters $q = (0.5, 3.0)$ is shown in Figure 6.3.2. However, the values of the parameters do not describe the quantity of physical interest in this application. The quantity we wish to compute is described by the forces acting on a part of the boundary $\Gamma_A \subset \Gamma_1$.

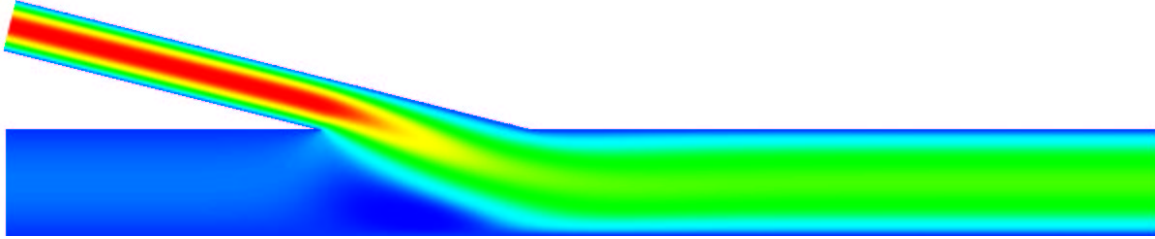


FIGURE 6.3.2. Solution of the state equation (horizontal velocity) for the exact parameters $q = (0.5, 3.0)$.

This quantity of interest is given by the functional E :

$$E(u) = c_0 \int_{\Gamma_A} n \cdot \sigma \cdot d \, ds, \quad (6.3.157)$$

where $d = (-1, 0)$ is a chosen direction, c_0 is a given constant, and σ denotes the stress tensor given as usual by:

$$\sigma = \frac{\nu}{2}(\nabla v + (\nabla v)^T) - pI.$$

For the evaluation of E we use the following representation:

$$E(u) = c_0(a(u)(\phi_d) - f(\phi_d)), \quad \phi_d = (\psi_d, 0),$$

where ψ_d is a prolongation of the vector d to the whole domain, which vanishes on $\partial\Omega \setminus \Gamma_A$. For the discussion of this representation including the independence of the prolongation and an a priori error analysis see Giles, Larson, Levenstam & Süli [53] and Becker [11].

The described problem is solved by the Gauß-Newton method with the initial guess $q_0 = (0, 0)$, which corresponds to the state variable $u = 0$. Application of the a posteriori error estimator for the quantity of interest, see Section 5.3, leads to the sequence of locally refined meshes, see in Figure 6.3.3.

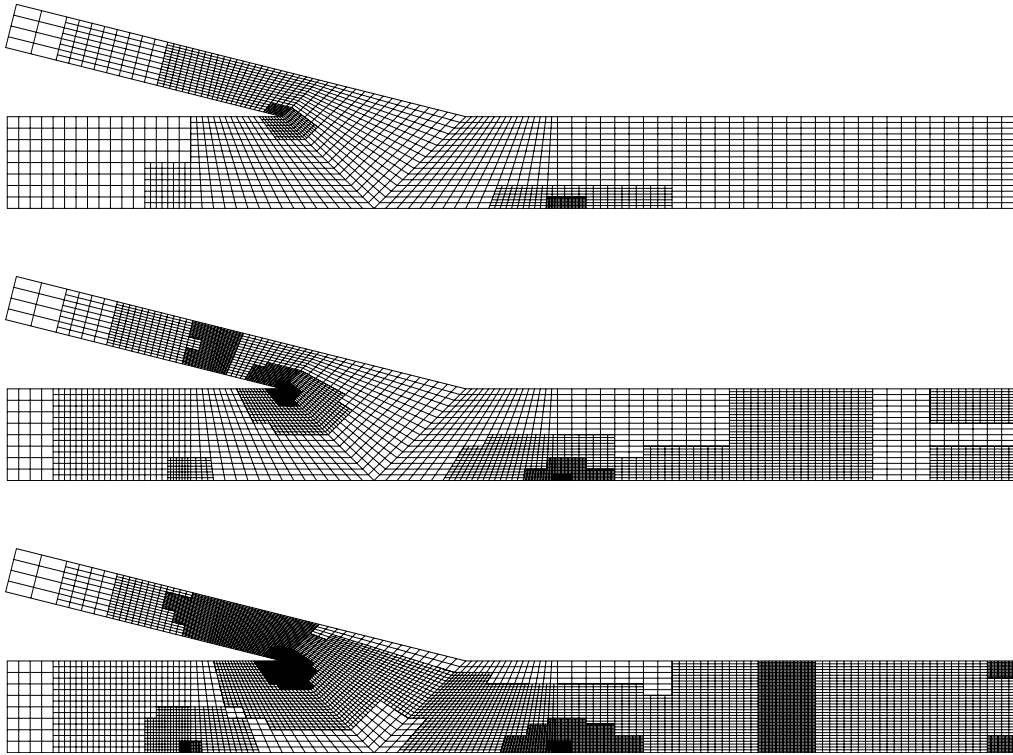


FIGURE 6.3.3. A sequence of the obtained meshes for computing the forces acting on Γ_A with 2 491, 5 297 and 10 029 nodes.

On the first (coarse) mesh 10 iterations of the optimization loop are done. However, on the following meshes only a few (≤ 3) iterations are required for sufficient reduction of the residual of the optimality condition, see the discretion of the optimization loop in Section 4.1.

The comparison of the error in the quantity of interest (6.3.157) for uniform mesh refinement, refinement based on an “energy” error estimator, and refinement produced by our a posteriori error estimator is done in Figure 6.3.4. It turns out, that the refinement strategy based on the error estimator for the quantity of interest leads to very efficient meshes. For achieving a tolerance $TOL = 1\%$ a local refined mesh with 5 297 nodes is needed. The usage of uniform refined meshes requires a mesh with 41 665 nodes for the same tolerance.

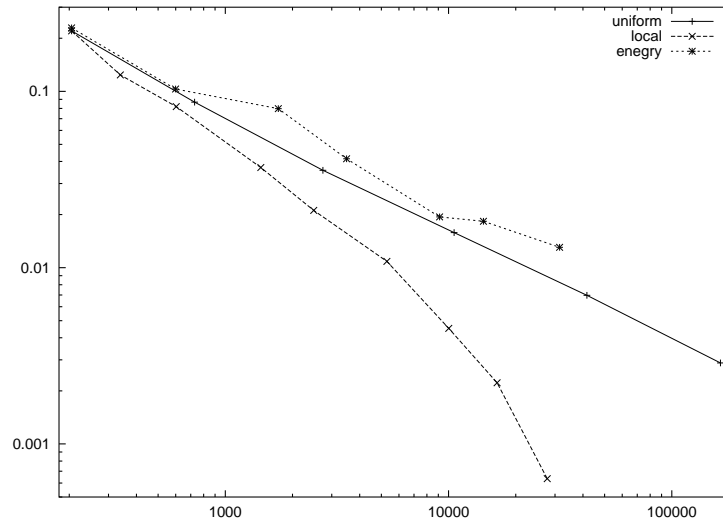


FIGURE 6.3.4. Error in the quantity of interest E vs. number of mesh points for different types of mesh refinement (uniform refinement, “energy”-based refinement, and local refinement resulting from our a posteriori error estimator).

6.4. Flow in a canal with a rough wall

In this section we discuss an example dealing with the estimation of parameters from Dirichlet boundary conditions. We consider a flow around a cylinder in a channel with a rough wall, see Figure 6.4.6. Due to the unknown roughness we wish to make the computation on a canal with even walls, see Figure 6.4.5.

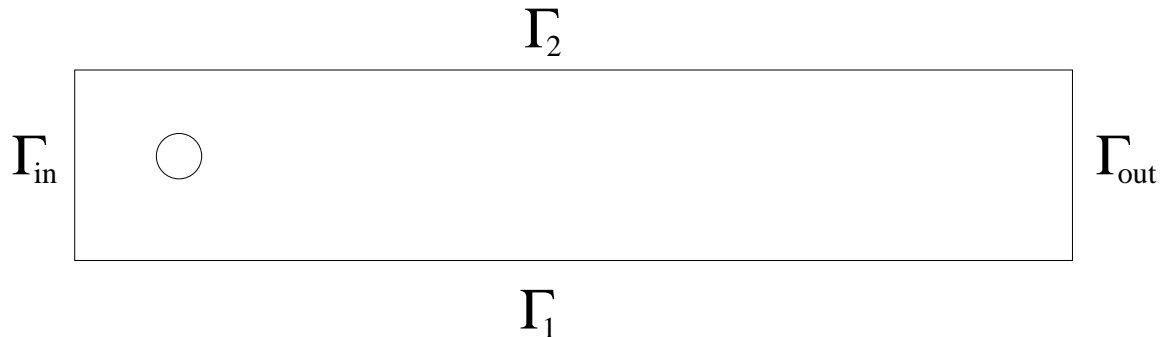


FIGURE 6.4.5. Computational domain with even walls.

This leads to unknown inhomogeneous Dirichlet boundary conditions. We parameterize these boundary conditions

$$v = g \text{ on } \Gamma_1$$

by means of Fourier sums with an additional compatibility condition $g(0) = 0$. This leads to the following representation for the components $g^{(1)}$ and $g^{(2)}$ of g :

$$g^{(k)} = \sum_{l=1}^L (q_{2l}^{(k)} \sin(l\omega x) + q_{2l+1}^{(k)} (\cos(l\omega x) - 1)), \quad k = 1, 2,$$

where $L \in \mathbb{N}$ and $\omega \in \mathbb{R}_+$ are given numbers. The parameters $q \in Q = \mathbb{R}^{4L}$ are estimated from point-wise measurements of the velocity and pressure. In the computation results below we use $L = 8$ and the measurements of velocity and pressure at 36 different points ξ_i . This corresponding cost functional is given by:

$$C_{3i-2}(u) = p(\xi_i), \quad C_{3i-1}(u) = v_1(\xi_i), \quad C_{3i}(u) = v_2(\xi_i), \quad i = 1, 2, \dots, 36.$$

The initial parameter values for the optimization loop are chosen $q_0 = 0$ describing the homogeneous boundary conditions, see Figure 6.4.7.

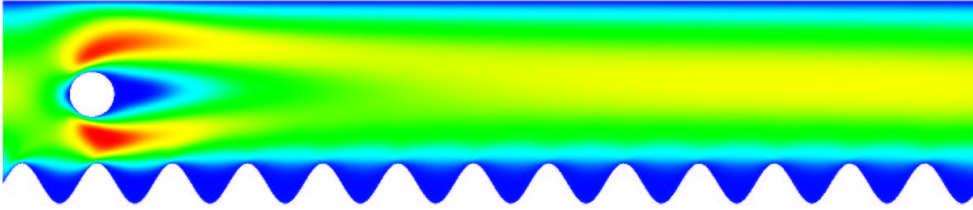


FIGURE 6.4.6. Flow around the cylinder computed on a canal with a rough wall.

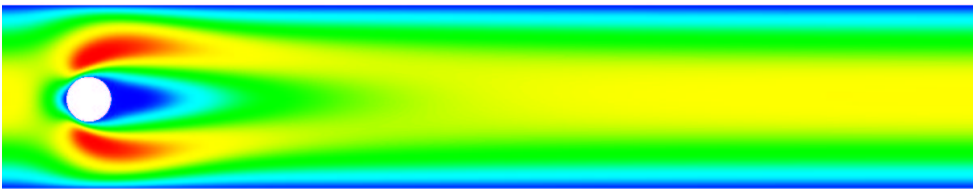


FIGURE 6.4.7. Flow around the cylinder computed on a canal with even walls and homogeneous Dirichlet boundary conditions.

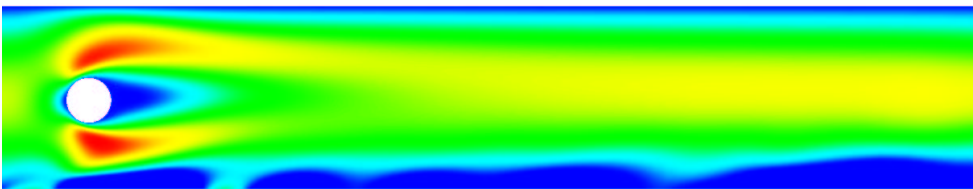


FIGURE 6.4.8. Flow around the cylinder with estimated boundary conditions.

The quantity of interest E is here the drag coefficient on the cylinder computed in the same way as in the previous section. The exact value for the drag coefficient is $E_{ref} \approx 6.6709$. The computation on the domain with even walls and homogeneous boundary conditions leads to the value $E(S(q_0)) \approx 5.5794$. However, using the estimated boundary conditions allows for obtaining an approximation $E(u) \approx 6.6755$, see Figure 6.4.8. A sequence of locally refined meshes produced for this computation is shown in Figure 6.4.9.

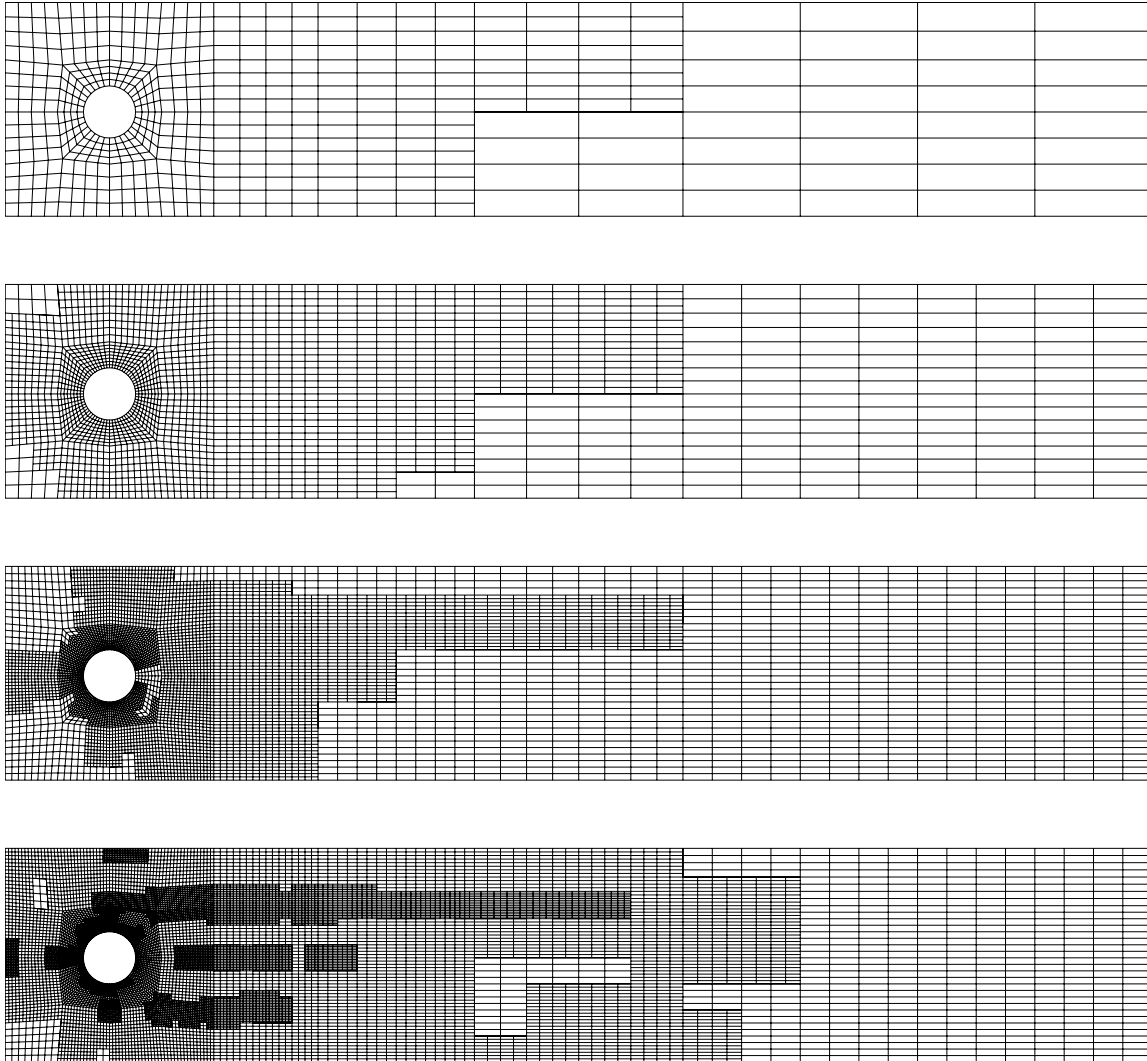


FIGURE 6.4.9. A sequence of the obtained meshes for computing the drag coefficient with estimated boundary conditions.

CHAPTER 7

Application to Multidimensional Reactive Flows

In this chapter we apply the algorithms described in previous chapters to two parameter identification problems for chemical models in multidimensional reactive flows. Due to heavy impact of physical and chemical models in combustion simulations the issue of parameter estimation is of essential importance for the development of accurate transport and chemical models. Typical problems are the estimation of Arrhenius parameters and diffusion coefficients. The examples considered in this chapter are introduced in Becker, Braack & Vexler [16]. The example in Section 7.1 deals with the estimation of Arrhenius parameters for a simple combustion model described by a convection-diffusion-reaction equation for one single reaction. In Section 7.2 we analyze diffusion parameters for a hydrogen flame. It includes the compressible Navier-Stokes equations and a system of convection-diffusion-reaction equations for nine chemical species and 38 elementary (bi-directional) reactions. To the author's knowledge, the paper [16] is the first published work on automatic parameter estimation for multidimensional computation of flames. For an overview of parameter estimation problems in chemistry, we refer to the book of Englezos and Kalogerakis [42]. Therein, many applications of parameter identification in the framework of ordinary differential equations are given. Parameter estimation problems for reactive flows in one space dimension are treated for instance by Bock et al. [90].

7.1. Identification of Arrhenius parameters

In this section we demonstrate the behavior of our methods for the problem of estimating Arrhenius parameters in a simple combustion model. The state equation is given by scalar stationary convection-diffusion-reaction equation for the variable u in a domain $\Omega \subset \mathbb{R}^2$ with a divergence-free vector field β and a diffusion coefficient D :

$$\beta \cdot \nabla u - \operatorname{div}(D \nabla u) + s(u, q) = f, \quad (7.1.158)$$

provided with Dirichlet boundary conditions $u = \hat{u}$ at the inflow boundary $\Gamma_{in} \subset \partial\Omega$ and Neumann conditions $\partial_n u = 0$ on $\partial\Omega \setminus \Gamma_{in}$. As usual in combustion problems, the reaction term is of Arrhenius type

$$s(u, q) := A \exp\{-E/(d - u)\}u(c - u). \quad (7.1.159)$$

The variable u stands for the mole fraction of a fuel, while the mole fraction of the oxidizer is $0.2 - u$. Since the Arrhenius law is a heuristic law and can not be derived by physical laws, the involved parameters are a priori unknown and have to be calibrated. This parameter fitting

is usually done by comparison of experimental data and simulation results. Therefore, this example is well suited for the proposed parameter identification algorithm. We assume the parameters d, c to be fixed and the parameters A, E are considered as unknown and form the vector-valued parameter $q = (\log(A), E) \in Q = \mathbb{R}^2$.

We consider the following configuration: Fuel (F) and oxidizer (Ox) are injected in different pipes and diffuse in a reaction chamber with overall length 35 mm and high 7 mm , see Figure 7.1.1. At the center tube, the Dirichlet condition for the fuel is $u = u_{in} := 0.2$, and at the upper and lower tube, $u = 0$. On all other parts of the boundary, homogeneous Neumann conditions are imposed.

The fix parameters in the Arrhenius law (7.1.159) are $c = u_{in}$ and $d = 0.24$. The convection direction $\beta(x, y)$ is a velocity field obtained by solving the incompressible Navier-Stokes equations with parabolic inflow profile at the tubes with peak flow $\beta_{max} = 0.2\text{ m/s}$. The diffusion coefficient D is chosen $D = 2.e-6$.

The measurements $C(u) \in Z = \mathbb{R}^{n_m}$ are modeled by mean values along $n_m = 10$ straight lines Γ_i at different positions in the reaction chamber, see dashed lines in Figure 7.1.1:

$$C_i(v) = \int_{\Gamma_i} v dx, \quad i = 1, \dots, n_m.$$

Instead of experimental data, we simply choose the optimal parameters to $q = (6.9, 0.07)$ and replace the measurements by computation with these parameters: $\bar{C} := C(S(q))$. As a consequence, the “measurements” perfectly match for the optimal parameters. This will not be the case in the example in Section 7.2.

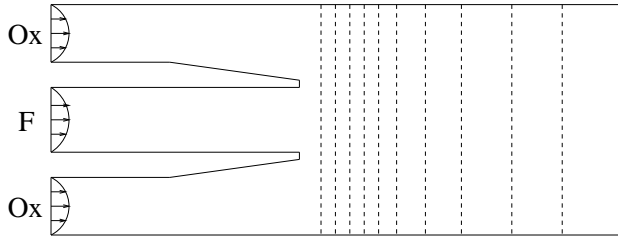


FIGURE 7.1.1. Configuration of the reaction chamber for estimating Arrhenius coefficients. Dashed vertical lines show the lines where the measurements are modeled.

The initial parameters are set to $q^0 = (\log(A^0), E^0) = (4, 0.15)$, leading to low reaction rates and a diffusion dominated solution. In Figure 7.1.2 the corresponding state variable (fuel) u^0 is shown. For the optimal q , in contrast to the initial guess q^0 , a sharp reaction front occurs, see Figure 7.1.3. Obviously, the difference in the parameters has a substantial impact to the state variable u .

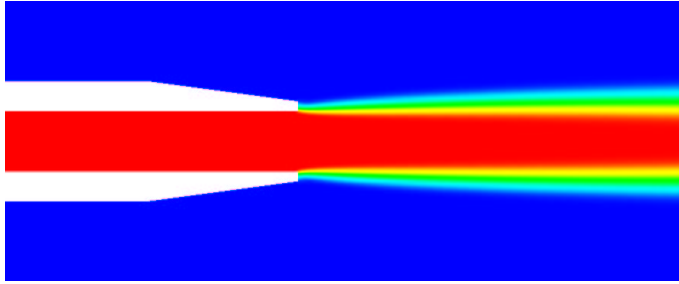


FIGURE 7.1.2. Mole fraction of the fuel (u^0) for the initial parameters q^0 . Blue: $u = 0$, red: $u = 0.2$.

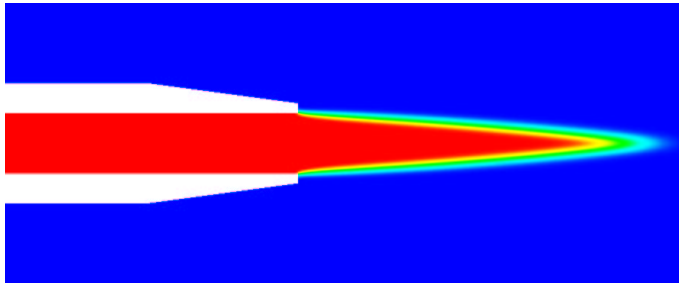


FIGURE 7.1.3. Mole fraction of the fuel (u) for the optimal parameters q .

The state equation (7.1.158) is discretized by finite elements described in Chapter 3. The stabilization of the convection term $\beta \cdot \nabla u$ is treated in the same way as in Section 6.1 for the Navier-Stokes equations. For the error functional, we choose the discretization error with respect to the second parameter, i.e. $E(q) = q_2$. In the optimization loop, we use the Gauß-Newton algorithm with the trust region strategy described in Section 4.3.

In Table 7.1.1, the results obtained are listed. The third column displays the norm of the least squares residual. On the first mesh with only $N = 1664$ nodes, 8 iterations (see second column) are done. On this mesh, the cost functional is reduced by more than two digits. In the fourth column, the remaining residual of the optimization condition (2.3.28) (in the discrete form) is listed:

$$Res := \|G_h^*(\bar{C} - c_h(q_h^k))\|. \quad (7.1.160)$$

The last two columns show the corresponding obtained parameters. After a substantial reduction of Res , the mesh is adapted locally according to the a posteriori error estimator, see Chapter 5. The second mesh has 2852 nodes. Here, the optimization loop is repeated. However on the finer meshes, only a few (≤ 3) iterations are necessary. On the finest mesh, the error in the first parameter is about 0.03% and in the second parameter about 0.3%.

TABLE 7.1.1. Numerical results for Arrhenius parameter identification.

N	it	$\ C(u) - \bar{C}\ $	Res	q_1	q_2
1664	1	5.87e-2	2.54e-3	4.000	1.500e-1
	2	5.86e-2	2.56e-3	4.001	1.499e-1
	3	5.81e-2	2.83e-3	4.132	1.499e-1
	4	4.47e-2	6.42e-3	5.630	1.489e-1
	5	2.36e-2	5.58e-3	6.752	1.481e-1
	6	6.22e-3	1.90e-3	7.433	1.093e-1
	7	8.34e-4	2.55e-4	6.660	4.621e-2
	8	3.00e-4	1.00e-4	6.825	6.394e-2
2852	1	4.79e-4	1.44e-4	6.798	6.216e-2
	2	3.68e-5	1.04e-5	6.905	7.134e-2
	3	1.92e-5	5.51e-8	6.906	7.158e-2
6704	1	2.33e-4	7.72e-5	6.906	7.158e-2
	2	1.42e-5	1.91e-8	6.904	7.066e-2
13676	1	6.91e-5	2.30e-5	6.904	7.063e-2
	2	3.53e-6	6.76e-9	6.905	7.052e-2
21752	1	1.22e-5	3.24e-6	6.905	7.052e-2
	2	2.84e-6	8.89e-9	6.902	7.022e-2

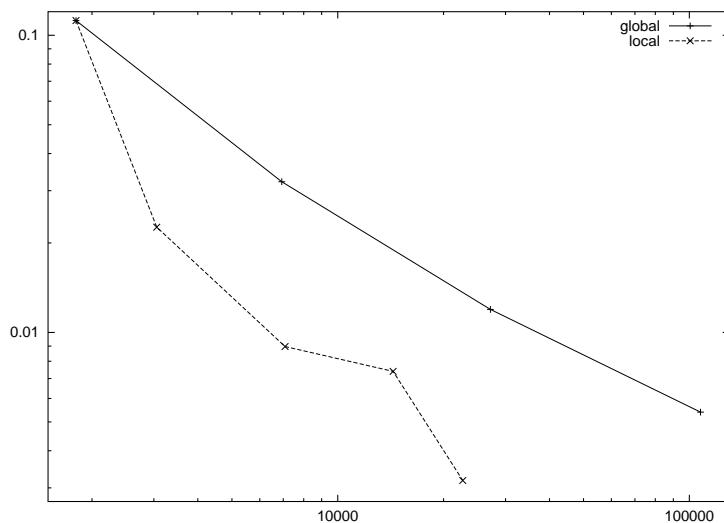


FIGURE 7.1.4. Relative error in the second Arrhenius parameter in dependence of the number of mesh points. Solid line: globally refined meshes, dashed line: locally refined meshes on the basis of a posteriori error estimation.

Comparing the error in the parameters with a more conventional strategy on globally refined meshes, our proposed algorithm is much more efficient. In Figure 7.1.4, the relative error in the second parameter is plotted in dependence of the number of mesh points. The dashed line results from our method on locally refined meshes. The solid line stands for parameter estimation with the same optimization loop but on uniformly refined meshes. For a relative error of less than 1%, only 6 704 nodes are necessary with a locally refined mesh, whereas more than 100 000 nodes are necessary on a uniformly refined mesh.

In Figure 7.1.5, a sequence of locally refined meshes produced by the refinement algorithm is shown. The highest amount of mesh points is located near the flame front and close to the measurement lines.

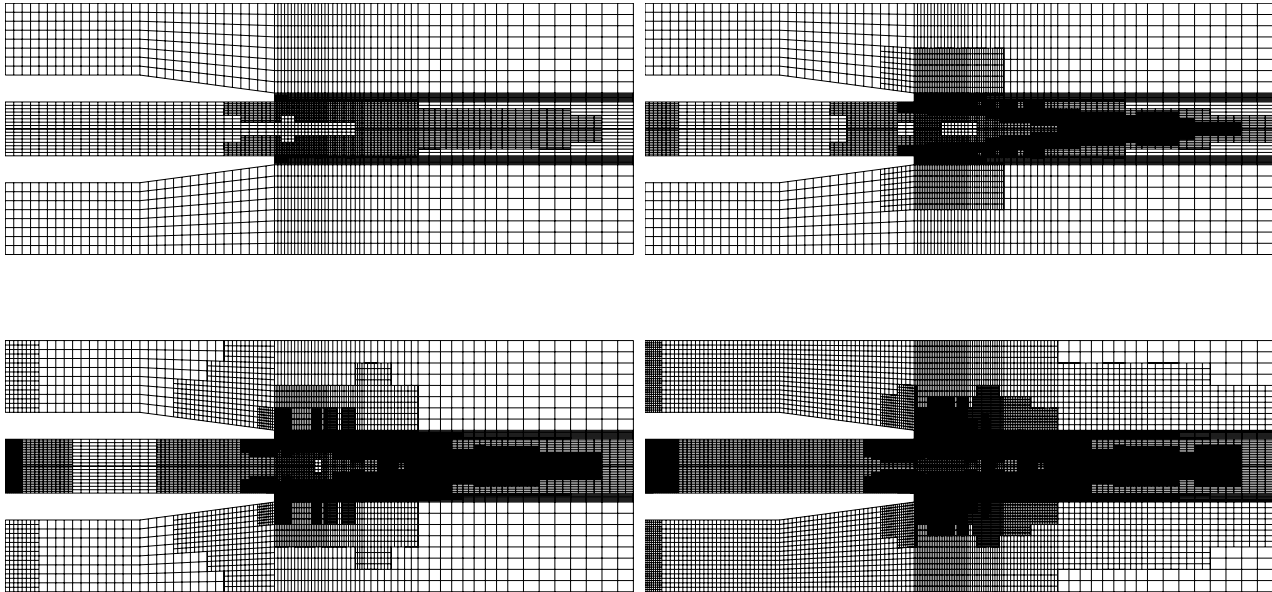


FIGURE 7.1.5. Obtained meshes for estimating Arrhenius parameters with 2 825, 6 704, 13 676 and 21 752 nodes (from upper left to lower right).

7.2. Identification of diffusion parameters

In this section we consider a stationary hydrogen diffusion flame and estimate parameters in a diffusion model. The propose is to calibrate the parameters in the Fick's law in order to substitute multicomponent diffusion laws. A similar problem for the ozone decomposition flame is discussed in Becker, Braack & Vexler [17].

7.2.1. UNDERLYING SYSTEM OF EQUATIONS

A stationary hydrogen diffusion flame considered here is modeled by the following system of equations for velocities v , pressure p , temperature T and mass fractions y_k :

$$\begin{aligned} \operatorname{div}(\rho v) &= 0, \\ (\rho v \cdot \nabla)v + \operatorname{div} \pi + \nabla p &= 0, \\ \rho v \cdot \nabla T - \frac{1}{c_p} \operatorname{div} \mathcal{Q} &= -\frac{1}{c_p} \sum_{i \in \mathcal{S}} h_i f_i, \\ \rho v \cdot \nabla y_k + \operatorname{div} \mathcal{F}_k &= f_k \quad k \in \mathcal{S}. \end{aligned}$$

The specific enthalpies are denoted by h_i , the heat capacity at constant pressure is denoted by c_p . Both quantities are evaluated by the use of thermodynamic data bases. The set \mathcal{S} denotes the set of chemical species. The density ρ is given by the perfect gas law in a mixture with partial molecular weights m_i and the uniform gas constant R :

$$\rho = \frac{p}{RT} \left(\sum_{i \in \mathcal{S}} \frac{y_i}{m_i} \right)^{-1}.$$

The stress tensor π is given as usual for compressible flows:

$$\pi = -\mu \left\{ \nabla v + (\nabla v)^T - \frac{2}{3}(\operatorname{div} v)I \right\},$$

where μ denotes the viscosity depending on temperature and the mixture fractions. The reaction terms f_i are modeled by a set of elementary reactions with reaction rates k_r of Arrhenius type:

$$\begin{aligned} f_i &= m_i \sum_{r \in \mathcal{R}} (\nu'_{ri} - \nu_{ri}) k_r \prod_{s \in \mathcal{S}} c_s^{\nu_{rs}}, \\ k_r &= A_r T^{\beta_r} \exp \left\{ -\frac{E_r}{RT} \right\}. \end{aligned}$$

The set \mathcal{R} includes all reactions considered. The stoichiometric coefficients of the products and educts for reaction r are denoted by ν'_{ri} and ν_{ri} , respectively. The concentration c_i of species i is given by $c_i = \rho y_i / m_i$. The heat flux \mathcal{Q} is given by the Fourier's law

$$\mathcal{Q} = -\lambda \nabla T,$$

where λ is the heat conductivity. The species fluxes \mathcal{F}_k are modeled by a simple Fick's law:

$$\mathcal{F}_k = q_k D_k^* \nabla y_k.$$

The scaling parameters q_k are the free parameters which have to be calibrated in the optimization procedure. Following Hirschfelder & Curtiss [66] and Warnatz, Dibble & Maas [88], the diffusion coefficients in the mixture D_k^* are given by

$$D_k^* = (1 - y_k) \left(\sum_{l \neq k} \frac{x_l}{D_{kl}^{bin}} \right)^{-1},$$

with binary diffusion coefficients D_{kl}^{bin} and mole fractions x_l . In order to insure that the sum over all species mass fractions sum up to 1 and to have a consistent model, the inert species (N_2) is erased from the set of unknown species.

For the system of equation described above we apply the approximation for Low-Mach number combustion due to Majda [74]. The discretization is done by bilinear finite elements, see Chapter 3. The stabilization of the pressure-velocity coupling and of the convection terms is treated in a similar way as in Section 6.1 for the Navier-Stokes equations. For the detailed description of the Low-Mach number approximations for combustion and the corresponding finite element discretization we refer to Braack [27] and Becker, Braack & Rannacher [15].

7.2.2. CONFIGURATION OF A HYDROGEN DIFFUSION FLAME

The set up for this hydrogen diffusion flame is taken from Braack & Ern [30] and is shown schematically in Figure 7.2.6. At the inflow boundary of the center tube, 10% mass fractions of hydrogen and 90% nitrogen is opposed. At the upper and lower tube, $y_{O_2} = 0.22$ and $Y_{N_2} = 0.78$ is prescribed. The peak velocity of the three parabolic velocity profiles is 1 m/s . Low temperature $T = 273 \text{ K}$ is prescribed at the inflow boundaries. On the other three boundaries (upper, lower, right), homogeneous Neumann conditions for temperature and mass fractions are opposed. For the velocities, the upper and lower wall are no-slip walls. The chemical model involves nine chemical species, see the mechanism in Table 7.2.2.

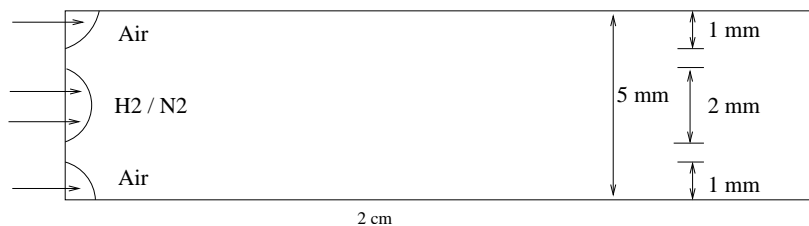


FIGURE 7.2.6. Schematic set up for the hydrogen diffusion flame.

TABLE 7.2.2. Reaction mechanism used for hydrogen combustion. The collision efficiencies are: $M(H, H_2, O_2, OH, H_2O, N_2) = (0, 1, 0.4, 0, 6.5, 0.4)$.

Reaction	A_r	β_r	E_{ar}
$O_2 + H=OH + O$	2.000×10^{14}	0	70.3
$H_2 + O=OH + H$	5.060×10^{04}	2.67	26.3
$H_2 + OH=H_2O + H$	1.000×10^{08}	1.6	13.8
$OH + OH=H_2O + O$	1.500×10^{09}	1.14	0.42
$H + H + M=H_2 + M$	1.800×10^{18}	-1	0
$O + O + M=O_2 + M$	2.900×10^{17}	-1	0
$H + OH + M=H_2O + M$	2.200×10^{22}	-2	0
$H + O_2 + M=HO_2 + M$	2.300×10^{18}	-0.8	0
$HO_2 + H=OH + OH$	1.500×10^{14}	0	4.2
$HO_2 + H=H_2 + O_2$	2.500×10^{13}	0	2.9
$HO_2 + H=H_2O + O$	3.000×10^{13}	0	7.2
$HO_2 + O=OH + O_2$	1.800×10^{13}	0	-1.7
$HO_2 + OH=H_2O + O_2$	6.000×10^{13}	0	0
$HO_2 + HO_2=H_2O_2 + O_2$	2.500×10^{11}	0	-5.2
$OH + OH + M=H_2O_2 + M$	3.250×10^{22}	-2	0
$H_2O_2 + H=H_2 + HO_2$	1.700×10^{12}	0	15.7
$H_2O_2 + H=H_2O + OH$	1.000×10^{13}	0	15
$H_2O_2 + O=OH + HO_2$	2.803×10^{13}	0	26.8
$H_2O_2 + OH=H_2O + HO_2$	5.400×10^{12}	0	4.2

The initial parameters are set to $q_0 = (1, \dots, 1) \in Q = \mathbb{R}^{n_p}$, $n_p = 8$, so that the Fick's law with conventional diffusion parameters is recovered. The right boundary is the natural outflow boundary condition. Due to the rich fuel condition at the center tube, there is not enough oxygen for a complete burning of the fuel. This kind of flame is called *under-ventilated*. However, the peak temperature reaches more than 2000 K in the burnt gas. In Figure 7.2.7, a zoom of the resulting mass fractions of hydrogen peroxide H_2O_2 is shown indicating the flame front.

We substitute the experimental data by a computation of the same flame but with a more sophisticated species diffusion model, namely multicomponent diffusion, see Ern & Giovangigli [46]. The diffusion fluxes are evaluated by the EGLIB library [45]. The corresponding flame is shown in Figure 7.2.8, showing a qualitatively different flame front. The heavy impact of diffusion models for hydrogen flames is also observed in Ern & Giovangigli [47]. We like to emphasize that this setting is not chosen to propose better constants for the Fick's laws but to show the capability of the parameter identification algorithm.

The vector of observation \bar{C} consists of 640 point values of mass fractions of hydrogen peroxide. These "observation points" are equidistant distributed in the region of the flame front $\Omega' = [0, 3.125 \text{ mm}] \times [0, 2.5 \text{ mm}]$, where we obtained reference values by computations with the multicomponent diffusion models. Unlike the example in Section 7.1, the observation

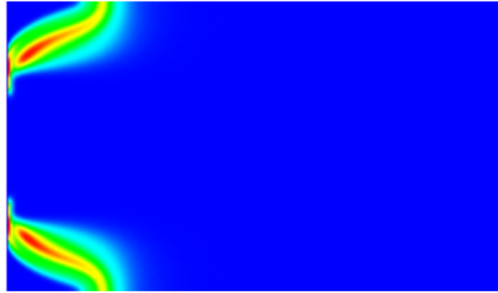


FIGURE 7.2.7. Mass fractions of H_2O_2 for the initial diffusion model (Fick's law).

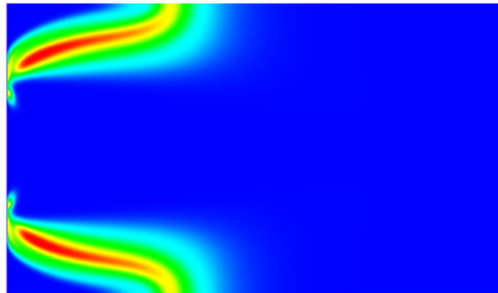


FIGURE 7.2.8. Mass fractions of H_2O_2 for the multicomponent diffusion model.

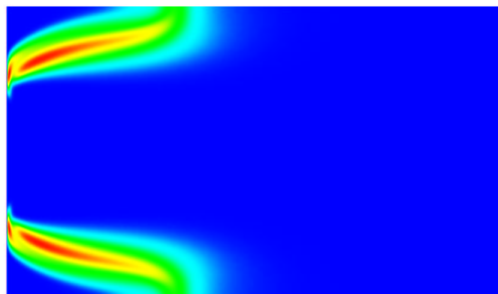


FIGURE 7.2.9. Mass fractions of H_2O_2 for the calibrated Fick's diffusion model.

operator for the optimized parameters does not exactly match with \bar{C} , because the Fick's law and multicomponent diffusion are qualitatively different.

7.2.3. COMPUTATIONAL RESULTS FOR THE HYDROGEN DIFFUSION FLAME

In the optimization loop we use the Gauß-Newton algorithm with the update formula for a part of the Hessian involving the second derivatives of the observation operator, as described in Section 4.2.3.

After optimization, the cost functional J is reduced by the factor of 5 and the obtained parameters are:

$$q = (0.269, 0.454, 1.09, 2.46, 0.946, 18.6, 0.579, 0.179).$$

A comparison of H_2O_2 for multicomponent diffusion in Figure 7.2.8 and the corresponding solutions with the parameter fitted Fick's law, given in Figure 7.2.9, shows a substantial improvement. The sequence of (zooms of) the locally refined meshes resulting from the application of the adaptive mesh refinement algorithm, described in Chapter 5, is shown in Figure 7.2.10.

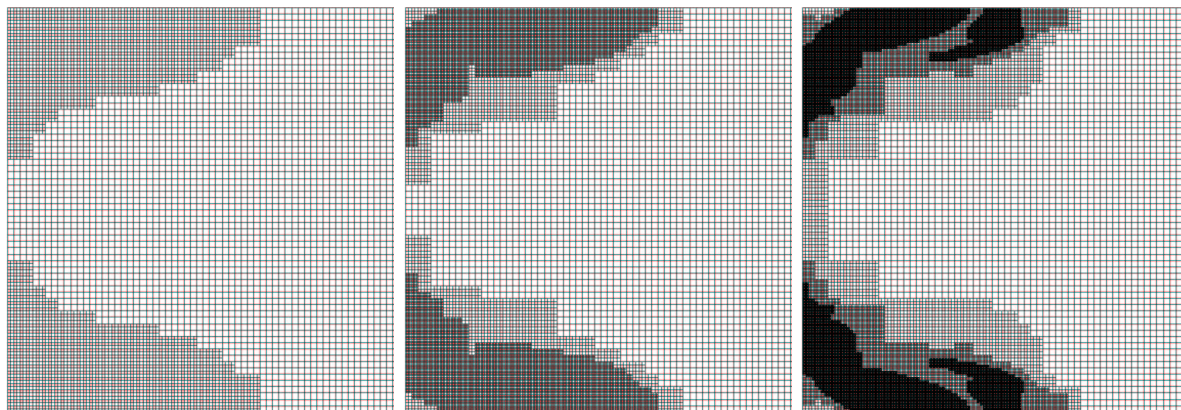


FIGURE 7.2.10. A sequence of zooms of the locally refined meshes for the hydrogen flame.

CHAPTER 8

Conclusion and Future Work

In this doctoral thesis efficient algorithms for parameter identification problems are developed and analyzed. Parameter identification problems governed by partial differential equations are discretized by finite element methods. For the error in parameters due to this discretization a priori and a posteriori analysis is developed. We show an optimal order of convergence for a class of parameter identification problems governed by elliptic partial differential equations of second order. Moreover, we discuss an extension of this proof for the case of point-wise measurements, which is important for applications.

The main issue of this work is the development of an adaptive mesh refinement algorithm with the aim of finding an efficient discretization for parameter identification. The algorithm produces a sequence of locally refined meshes based on the a posteriori error estimator for the error in parameters. The presented error estimator is derived in the general nonlinear case. Exploiting the special structure of parameter identification problem makes the computation of the error estimator cheap in comparison to the overall optimization loop. The quality of the resulting meshes and the efficiency of error estimation is illustrated for several example problems.

Different Newton type optimization algorithms acting on the parameter space are discussed and compared. Trust-region techniques are used for globalization of convergence. The underlying nonlinear state equations are solved by the Newton's method and the solutions of the linear subproblems are computed using standard multigrid algorithms. All computations are done on the basis of the package *ParamGascoigne* for treating parameter identification problems. This C++ library is implemented by the author as an extension of the finite element toolkit *Gascoigne3D*, see [12].

In order to demonstrate the capability of the presented approach we apply the developed methods to parameter identification problems in fluid dynamics and to estimation of chemical models in multidimensional reactive flow problems. In the context of CFD problems we discuss the treatment of unknown Dirichlet and Neumann boundary conditions. This is done by parameterization of the boundary conditions leading to parameter identification problems. Another application field is the calibration of the parameters in Arrhenius law and of the diffusion coefficients by modeling combustion processes with detailed chemistry. To the author's

knowledge, this is the first published result discussing automatic parameter estimation for multidimensional computation of flames.

An extension of the approaches presented in this thesis to estimating parameters from infinite dimensional spaces is a subject of future work. Here, we wish to combine model reduction techniques with adaptive mesh refinement for efficient handling of parameter identification problems with distributed parameters. Moreover, a posteriori error analysis may be also used for control of the regularization terms from the cost functional.

Another generalization of our work is treatment of time-dependent parameter identification problems. An appropriate extension of our mesh refinement techniques requires a posteriori analysis of the influence of the time and space discretizations on the quality of the estimated parameters. Another important issue is the data handling for the solution of the time-dependent adjoint problems.

Further, we intent to continue the work on the parameter estimation in combustion problems. Here, we wish to apply the proposed concepts to parameter identification with real experimental data.

Bibliography

- [1] M. Ainsworth and J.T. Oden. A unified approach to a posteriori error estimation using element residual methods, *Numer. Math.*, 65:23-50, 1993.
- [2] H.W. Alt. *Lineare Funktionalanalysis*, Springer Berlin, 1999.
- [3] N. Arada, E. Casas and F. Tröltzsch. Error estimates for the numerical approximation of a semilinear elliptic control problem, *J. Comput. Optim. Appl.* 23, No.2, 201-229, 2002.
- [4] I. Babuška and W.C. Rheinboldt. Error estimates for adaptive finite element computations, *SIAM J. Numer. Anal.*, 15:736-754, 1978.
- [5] I. Babuška and W.C. Rheinboldt. A posteriori error estimates for the finite element method, *Int. J. Numer. Meth. Engrg*, 12:1597-1615, 1978.
- [6] I. Babuška and A.D. Miller. A feedback finite element method with a posteriori error estimation, *Comput. Methods Appl. Mech. Engrg*, 61:1-40, 1987.
- [7] W. Bangerth. Adaptive finite element methods for the identification of distributed coefficient in partial differential equations, PhD thesis, University of Heidelberg, 2002.
- [8] R.E. Bank and A. Weiser. Some a posteriori error estimators for elliptic partial differential equations, *Math. Comp.*, 44:283-301, 1985.
- [9] H.T. Banks and K. Kunisch. *Estimation techniques for distributed parameter systems*, Boston: Birkhäuser, 1989.
- [10] D. Bates and D. Watts. *Nonlinear regression analysis and its applications*, John Wiley & Sons, New York, 1988.
- [11] R. Becker. Adaptive finite elements for optimal control problems, Habilitationsschrift, Institut für Angewandte Mathematik, Universität Heidelberg, 2001.
- [12] R. Becker and M. Braack. Gascoigne3D, A finite element toolkit for flow problems, SFB-preprint, Heidelberg, 2003.
- [13] R. Becker and M. Braack. Multigrid techniques for finite elements on locally refined meshes, *Numerical Linear Algebra with Applications*, 7:363-379, 2000. Special Issue.
- [14] R. Becker and M. Braack. A finite element pressure gradient stabilization for the Stokes equations based on local projections, *Calcolo* 38, No.4, 137-199, 2001.
- [15] R. Becker, M. Braack and R. Rannacher. Numerical Simulation of Laminar Flames at Low Mach Number with Adaptive Finite Elements, *Combustion Theory and Modelling*, Bristol, Vol. 3, Nr. 3, p. 503-534, 1999.
- [16] R. Becker, M. Braack and B. Vexler. Numerical parameter estimation for chemical models in multidimensional reactive flows, *Combustion Theory and Modelling*, submitted, 2003.
- [17] R. Becker, M. Braack and B. Vexler. Parameter identification for chemical models in combustion problems, in *Proc. of Chemnitz FEM Symposium*, Sep. 22-24, submitted, 2003.
- [18] R. Becker and R. Rannacher. 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.
- [19] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods, In *Acta Numerica 2001* (A. Iserles, ed.), Cambridge University Press, Cambridge, 1-102, 2001.

- [20] R. Becker, H. Kapp, and R. Rannacher Adaptive finite element methods for optimal control of partial differential equations: Basic concept, *SIAM J. Cont. Opt.***39**(1), 113-132, 2000.
- [21] R. Becker and B. Vexler. Mesh Adaptation for Parameter Identification Problems, in Proc. of the 8th Conference on Control of Distributed Parameter Systems Graz, July 15-21, 2001, Intern. Series of Numerical Mathematics, Vol. 143, 41-56, 2002.
- [22] R. Becker and B. Vexler. A posteriori error estimation for finite element discretization of parameter identification problems, *Numerische Mathematik*, Vol. 96-3, 435-459, 2004.
- [23] R. Becker and B. Vexler. Calibration of PDE Models and Sensitivity Analysis with Adaptive Finite Elements: Application to CFD Problems, *Journal of Computational Physics*, submitted, 2003.
- [24] H.G. Bock. Recent Advances in Parameter Identification Techniques for O.D.E., *Progress in Scientific Computing* (Deuffhard, Hairer, eds.), Birkhäuser, Boston, 1983.
- [25] H.G. Bock. Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen, *Bonner Mathematische Schriften*, 1987.
- [26] H.G. Bock, E. Eich and J. Schlöder. Numerical solution of constrained least squares boundary value problems in differential-algebraic equations, In K. Strehmel, *Numerical Treatment of Differential Equations*, BG Teubner, Leipzig, 1988.
- [27] M. Braack. An Adaptive Finite Element Method for Reactive Flow Problems, Dissertation, Universität Heidelberg, 1998.
- [28] M. Braack. Localization of error estimators, SFB-preprint, Heidelberg, in preparation, 2003.
- [29] M. Braack and A. Ern. A posteriori control of modeling errors, *SIAM J. Multiscale Modeling and Simulation*, 1(2):221-238, 2003.
- [30] M. Braack and A. Ern. Coupling multimodeling with local mesh refinement for the numerical solution of laminar flames, *Combust. Theory Modelling*, submitted, 2003.
- [31] D. Braess. *Finite Elemente. Theorie, schnelle Löser und Anwendungen in der Elastizitätstheorie*, Springer, Berlin, 2003.
- [32] S. Brenner and R.L. Scott. *The mathematical theory of finite element methods*, Springer, Berlin Heidelberg New York, 1994.
- [33] C.F. Carey and J.T. Oden. *Finite Elements, Computational Aspects, Vol. III.*, New Jersey, Prentice-Hall, 1984.
- [34] P.G. Ciarlet. *The Finite Element Method for Elliptic Problems*, North-Holland Publishing Company, Amsterdam, 1978.
- [35] Ph. Clément. Approximation by finite element functions using local regularization, *Revue Franc. Automat. Inform. Rech. Operat.* 9(R-2), 77-84, 1975.
- [36] A.R. Conn, N. Gould and Ph.L. Toint. *Trust-region methods*, SIAM, MPS, Philadelphia, 2000.
- [37] K. Deckelnick and M. Hinze. Error estimates in space and time for tracking-type control of the instationary Stokes system, in Proc. of the 8th Conference on Control of Distributed Parameter Systems Graz, July 15-21, 2001, Intern. Series of Numerical Mathematics, Vol. 143, 87-104, 2002.
- [38] J.E. Dennis and R.B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, Number 16 in *Classics in Applied Mathematics*. Philadelphia, PA: SIAM, Society for Industrial and Applied Mathematics, 1996.
- [39] J.E. Dennis, D.M. Gay and R.E. Welsch. An adaptive nonlinear least-squares algorithm, [J] *ACM Trans. Math. Softw.* 7, 348-368, 1981.
- [40] J. Dieudonné. *Fondation of Modern Analysis*, Academic Press, New York, 1960.
- [41] P. Deuffhard, H.W. Engl and O. Scherzer. A convergence analysis of iterative methods for the solution of nonlinear ill-posed problems under affinity invariant conditions, *Inverse Probl.* 14, No.5, 1081-1106, 1998.
- [42] P. Englezos and N. Kalogerakis. *Applied parameter estimation for chemical engineers*, Marcel Dekker, New York, Basel, 2001.

- [43] K. Eriksson and C. Johnson. An adaptive finite element method for linear elliptic problems, *Math. Comp.*, 50:361-383, 1988.
- [44] K. Eriksson and C. Johnson. Adaptive finite element method for parabolic problems, I: A linear model problem, *SIAM J. Numer. Anal.*, 28:43-77, 1991.
- [45] A. Ern and V. Giovangigli. <http://www.cmap.polytechnique.fr/www.eglib>.
- [46] A. Ern and V. Giovangigli. *Multicomponent Transport Algorithms*, Lecture Notes in Physics, m24, Springer, 1994.
- [47] A. Ern and V. Giovangigli. Thermal diffusion effects in hydrogen-air and methane-air flames, *Combust. Theory Modelling*, 2:349-372, 1998.
- [48] R. S. Falk. Approximation of a class of optimal control problems with order of convergence estimates, *J. Math. Anal. Appl.* 44, 28-47, 1973.
- [49] R. S. Falk. Error Estimates of the Numerical Identification of a Variable Coefficient, *Math. Comp.*, 40(162):537-546, 1983.
- [50] R. Fletcher. *Practical methods of optimization*, John Wiley, Chichester, 2001.
- [51] J. Frehse and R. Rannacher. Eine L^1 -Fehlerabschätzung für discrete Grundlösungen in der Methode der finiten Elemente, *Tagungsband Finite Elemente*, Bonner Math. Schriften, 89:92-114, 1976.
- [52] A.V. Fursikov. *Optimal control of distributed systems. Theory and applications*, vol. 187 of *Translations of Mathematical Monographs*. American Mathematical Society, 2000.
- [53] M. Giles, M. Larson, M. Levenstam and E. Suli. Adaptive error control for finite element approximations of the lift and drag coefficients in viscous flow, *Technical Report NA-76/06*, Oxford University Computing Laboratory, 1997.
- [54] V. Girault and P.-A. Raviart. *Finite Elements for the Navier Stokes Equations*, Springer, Berlin, 1986.
- [55] R. Griesse. *Parametric Sensitivity Analysis in Optimal Control of a Reaction Diffusion System - Part I: Solution Differentiability*, Preprint No. 269, Special Research Center F 003, Graz, 2003.
- [56] R. Griesse. *Parametric Sensitivity Analysis in Optimal Control of a Reaction Diffusion System - Part I: Practical Methods and Examples*, Preprint No. 271, Special Research Center F 003, Graz, 2003.
- [57] P. Grisvard. *Singularities in Boundary Value Problems*, Masson, Paris, and Springer, Berlin, 1992.
- [58] Ch. Großmann and H.-G. Roos. *Numerik partieller Differentialgleichungen*, B. G. Teubner, Stuttgart, 1994.
- [59] J.-L. Guermond. Stabilization of Galerkin approximations of transport equations by subgrid modeling, *Modél. Math. Anal. Numér.*, 33(6):1293-1316, 1999.
- [60] M. Gunzburger and L. Hou. Finite-dimensional approximation of a class of constrained nonlinear optimal control problems, *SIAM J. Control Optimization* 34, No.3, 1001-1043, 1996.
- [61] E. Haber, U. Ascher and D. Oldenburg. On optimization techniques for solving nonlinear inverse problems, *J. Inverse Probl.* 16, No.5, 1263-1280, 2000.
- [62] G. Hellwig. *Partial differential equations*, B.G. Teubner, Stuttgart, 1977.
- [63] J.G. Heywood, R. Rannacher and S. Turek. Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations, *Int. J. Numer. Math. Fluids*, 22:325-352, 1992.
- [64] M. Hinze. *Optimal and instantaneous control of the instationary Navier-Stokes equations*, Habilitation thesis, Fachbereich Mathematik, Technische Universität Berlin, 2000.
- [65] M. Hinze and K. Kunisch. Second order methods for optimal control of time-dependent fluid flow, *SIAM J. Control Optimization* 40, No.3, 925-946, 2001.
- [66] J. O. Hirschfelder and C. F. Curtiss. *Flame and Explosion Phenomena*, Williams and Wilkins Cp., Baltimore, 1949.
- [67] K. M. S. Humak. *Nonlinear regression, functional relations and robust methods. Statistical methods of model building*, In H. Bunke und O. Bunke, ed., *Wiley Series in Probability and Mathematical Statistics*, John Wiley & Sons, Berlin, 1989.

- [68] C. Johnson. Numerical solution of partial differential equations by finite element method, Cambridge University Press, Cambridge, 1987.
- [69] T. Kärkkäinen. Error estimates for distributed parameter identification in linear elliptic equations, *J. Math. Syst. Estim. Control* 6, No.1, 117-120, 1996.
- [70] C. Kravaris and J.H. Seinfeld. Identification of parameters in distributed parameter systems by regularization, *SIAM J. Control Optim.*, 23:217-241, 1985.
- [71] J.L. Lions. Optimal control of systems governed by partial differential equations, Num. 170 in *Die Grundlehren der mathematischen Wissenschaften*. Springer Berlin, 1971.
- [72] V.G. Litvinov. Optimization in elliptic problems with applications to mechanics of deformable bodies and fluid mechanics, *Operator Theory: Advances and Applications*. 119. Basel: Birkhäuser, 2000.
- [73] W. Liu and N. Yan. A posteriori error estimates for convex boundary control problems, *SIAM J. Numer. Anal.*, 39(1):73–99, 2001.
- [74] A. Majda. *Compressible Fluid Flow and Systems of Conservation Laws in Several Space Variables*, Springer, New York, 1984.
- [75] K. Malanowski. Sensitivity analysis for parametric optimal control of semilinear parabolic equations, *J. Convex Anal.* 9, No.2, 543-561, 2002.
- [76] P. Neittaanmäki and D. Tiba. Optimal control of nonlinear parabolic systems: theory, algorithms, and applications, vol. 179 of *Pure and Applied Mathematics*. Marcel Dekker, 1994.
- [77] P. Neittaanmäki and X.-C. Tai. Error estimates for numerical identification of distributed parameters, *J. Comput. Math.* 10, Suppl., 66-78, 1992.
- [78] J. Nocedal and S.J. Wright. *Numerical Optimization*, Springer Series in Operations Research, Springer New York, 1999.
- [79] R. Rannacher. L^∞ -Stability estimates and asymptotic error expansion for parabolic finite element equations, in Tagungsband “Extrapolation and defect correction”, *Bonner Mathematische Schriften* 228, 74-94, 1991.
- [80] R. Rannacher. On Chorin’s projection method for the incompressible Navier-Stokes equations, in *Proc. of the Conference on the Navier-Stokes equation II*, held in Oberwolfach, August 18-24, 1991.
- [81] R. Rannacher and B. Vexler. A priori error analysis for finite element approximation of parameter identification problems, in preparation, 2003.
- [82] F. Riesz and B. Sz.-Nagy. *Vorlesungen über Funktionalanalysis*, VEB Deutscher Verlag der Wissenschaften Berlin, 1982.
- [83] G. A. F. Seber and A. J. Lee. *Linear regression analysis*, Wiley Series in Probability and Statistics, NJ: Wiley, Hoboken, 2003.
- [84] J. Stoer and R. Bulirsch. *Introduction to Numerical Analysis*, Texts in Applied Mathematics, Springer New York, 2002.
- [85] V.M. Tikhomirov. *Fundamental principles of the theory of extremal problems*, Wiley, 1986.
- [86] B. Vexler. A posteriori Fehlerschätzung und Gitteradaption bei Finite-Elemente-Approximationen nicht-linearer elliptischer Differentialgleichungen, Diplomarbeit, Universität Heidelberg, 2000.
- [87] B. Vexler. Adaptive finite elements for output-oriented model calibration, in *Proc. of International Conference on High Performance Scientific Computing*, Hanoi, March 10-14, 2003, submitted.
- [88] J. Warnatz, U. Maas, and R. W. Dibble. *Combustion*, Springer, New York, 3. edition, 2001.
- [89] E. Zeidler. *Nonlinear functional analysis and its applications I-IV*, Springer Berlin, 1985-1990.
- [90] M. W. Ziesse, H. G. Bock, J. V. Gallitzendörfer, and J. P. Schlöder. Parameter estimation in multispecies transport reaction systems using parallel algorithms, In J. Gottlieb and P. DuChateau, editors, *Parameter Identification and Inverse Problems in Hydrology, Geology and Ecology*, Kluwer academic publishers, 1996.