Acta Numerica (2001), pp. 1–102 © Cambridge University Press, 2001

An optimal control approach to *a posteriori* error estimation in finite element methods

Roland Becker and Rolf Rannacher
Institut für Angewandte Mathematik
Universität Heidelberg
INF 293/294, D-69120 Heidelberg, Germany
http://gaia.iwr.uni-heidelberg.de
E-mail: Roland.Becker@iwr.uni-heidelberg.de
Rolf.Rannacher@iwr.uni-heidelberg.de

This article surveys a general approach to error control and adaptive mesh design in Galerkin finite element methods that is based on duality principles as used in optimal control. Most of the existing work on *a posteriori* error analysis deals with error estimation in global norms like the 'energy norm' or the L^2 norm, involving usually unknown 'stability constants'. However, in most applications, the error in a global norm does not provide useful bounds for the errors in the quantities of real physical interest. Further, their sensitivity to local error sources is not properly represented by global stability constants. These deficiencies are overcome by employing duality techniques, as is common in *a priori* error analysis of finite element methods, and replacing the global stability constants by computationally obtained local sensitivity factors. Combining this with Galerkin orthogonality, *a posteriori* estimates can be derived directly for the error in the target quantity. In these estimates local residuals of the computed solution are multiplied by weights which measure the dependence of the error on the local residuals. Those, in turn, can be controlled by locally refining or coarsening the computational mesh. The weights are obtained by approximately solving a linear adjoint problem. The resulting *a posteriori* error estimates provide the basis of a feedback process for successively constructing economical meshes and corresponding error bounds tailored to the particular goal of the computation. This approach, called the 'dual-weighted-residual method', is introduced initially within an abstract functional analytic setting, and is then developed in detail for several model situations featuring the characteristic properties of elliptic, parabolic and hyperbolic problems. After having discussed the basic properties of duality-based adaptivity, we demonstrate the potential of this approach by presenting a selection of results obtained for practical test cases. These include problems from viscous fluid flow, chemically reactive flow, elasto-plasticity, radiative transfer, and optimal control. Throughout the paper, open theoretical and practical problems are stated together with references to the relevant literature.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

CONTENTS

1	Introduction	2
2	A paradigm for *a posteriori* error control	7
3	The dual-weighted residual method	17
4	Norm-based error estimation	32
5	Evaluation of estimates and mesh adaptation	39
6	Algorithmic aspects for nonlinear problems	43
7	Realization for time-dependent problems	48
8	Application to incompressible viscous flow	59
9	Application to thermal and reactive flow	66
10	Application to elasto-plasticity	76
11	Application to radiative transfer	82
12	Application to optimal control	87
13	Conclusion and outlook	92
References		95

1. Introduction

Solving complex systems of partial differential equations by discretization methods may be considered in the context of 'model reduction': a conceptually *infinite-dimensional* model is approximated by a *finite-dimensional* one. Here, the quality of the approximation depends on the proper choice of the discretization parameters, for example, the mesh width, the polynomial degree of the trial functions, and the size of certain stabilization parameters. As the result of the computation, we obtain an approximation to the desired output quantity of the simulation and, besides that, certain accuracy indicators, such as local cell residuals. Controlling the error in such an approximation of a continuous model of a physical system requires us to determine the influence factors of the local error indicators on the target quantity. Such a sensitivity analysis with respect to local perturbations of the model is common in optimal control theory and naturally introduces the concept of an 'adjoint' (or 'dual') problem.

For illustration, consider a *continuous* model governed by a linear differential operator A and a force term f, and a related discrete model depending on a discretization parameter $h \in \mathbb{R}_+$:

$$Au = f, \qquad A_h u_h = f_h. \tag{1.1}$$

In designing this discretization, we have to detect the interplay of the various error propagation effects in order to achieve

(i) *a posteriori error control*, that is, control of the error in quantities of physical interest such as stress values, mean fluxes, or drag and lift coefficients, and

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

(ii) *solution-adapted meshing*, that is, design of economical meshes for computing these quantities with optimal efficiency.

Traditionally, *a posteriori* error estimation in Galerkin finite element methods is done with respect to a natural 'energy norm' $\|\cdot\|_E$ induced by the underlying differential operator. This results in estimates of the form

$$\|u - u_h\|_E \leq c_s \|\rho(u_h)\|_E^*, \tag{1.2}$$

with a suitable dual norm $\|\cdot\|_E^*$ and the computable 'residual' $\rho(u_h) = f - Au_h$, which is well defined in the context of a Galerkin finite element method. Now, the main goal is to localize the residual norm in order to make it computable as a sum of cell-wise contributions. This approach was initiated by the pioneering work of Babuška and Rheinboldt (1978*a*, 1978*b*) and was then further developed by Ladeveze and Leguillon (1983), Bank and Weiser (1985), and Babuška and Miller (1987), to mention only a few of the most influential papers. For discussions and further references, we refer to the surveys by Verfürth (1996) and Ainsworth and Oden (1997*a*). In this article, 'energy error' estimation will be addressed only when it is important for our own topic.

Energy error estimation seems rather generic as it is directly based on the variational formulation of the problem and allows us to exploit its natural coercivity properties. However, in most applications the error in the energy norm does not provide useful bounds on the errors in the quantities of real physical interest. A more versatile method for *a posteriori* error estimation with respect to relevant error measures is obtained by using duality arguments as is common from the *a priori* error analysis of finite element methods (the so-called 'Aubin–Nitsche trick'). Let $J(u)$ be a quantity of physical interest derived from the solution u by applying a functional $J(\cdot)$. The goal is to control the error $J(u) - J(u_h)$ in terms of local residuals $\rho_K(u_h)$ computable on each of the mesh cells K. An example is control of the local total error $e_K = (u - u_h)_{|K}$. By superposition, e_K splits into two components, the locally produced 'truncation error' and the globally transported 'pollution error', $e_K^{\text{tot}} = e_K^{\text{loc}} + e_K^{\text{trans}}$, assuming for simplicity that the underlying problem is linear. In view of the error equation $A(u–u_h) = \rho(u_h)$, the effect of the cell residual ρ_K on the error $e_{K'}$, at another cell K', is governed by the Green function of the continuous problem. In practice it is mostly impossible to determine the complex error interaction by analytical means; rather, it has to be detected by computation. This leads to 'weighted' *a posteriori* error estimates

$$|J(u) - J(u_h)| \approx \langle \rho(u_h), \omega_h(z) \rangle, \tag{1.3}$$

where the sensitivity factor $\omega_h(z)$ is obtained by approximately solving an 'adjoint problem' $A^* z = j$, with j a density function associated with $J(\cdot)$.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

The adjoint solution z may be viewed as a generalized Green's function with respect to the output functional $J(\cdot)$, and accordingly the weight $\omega_h(z)$ describes the effect of local variations of the residual $\rho(u_h)$ on the error quantity $J(u)-J(u_h)$, for instance as the consequence of mesh adaptation. This approach to *a posteriori* error estimation is called the 'dual-weighted residual method' (or, for short, 'DWR method'). On the basis of *a posteriori* error estimates like (1.3), we can design a feedback process in which error estimation and mesh adaptation go hand-in-hand, leading to economical discretization for computing the quantities of interest. This approach is particularly designed for achieving high accuracy at minimum computational cost. The additional work required by the evaluation of the error bounds is usually acceptable since, particularly in nonlinear cases, it amounts to only a moderate fraction of the total costs.

The use of duality arguments in *a posteriori* error estimation goes back to ideas of Babuška and Miller (1984*a*, 1984*b*, 1984*c*) in the context of post-processing of 'quantities of physical interest' in elliptic model problems. It has since been systematically pursued by Eriksson and Johnson (1988, 1991) and their collaborators for more general situations (see also Johnson (1994) and the survey paper by Eriksson, Estep, Hansbo and Johnson (1995)). Here, stability constants for adjoint problems are mostly derived by analytical arguments. In Becker and Rannacher (1996*a*, 1996*b*) this approach is further developed into a computation-based feedback method, the 'DWR method', for error control and mesh optimization in computing local quantities of interest (for surveys and applications to problems in mechanics, physics and chemistry see Rannacher (1998*a*, 2000)). More recently, following ideas by Babuška and Miller (1984*a*, 1984*b*, 1984*c*), related techniques based on 'energy-norm' error estimation have been proposed by Machiels, Patera and Peraire (1998) and Oden and Prudhomme (1999).

We illustrate the ideas underlying the DWR method within an abstract setting. Let $A(\cdot,\cdot)$ be a bilinear form and $F(\cdot)$ a linear functional defined on some function space V, such that the given equation $Au = f$ has the variational formulation

$$A(u,\varphi) = F(\varphi), \quad \text{for all } \varphi \in V. \tag{1.4}$$

For a *finite-dimensional* subspace $V_h \subset V$, with a mesh-size parameter $h \in \mathbb{R}_+$, the Galerkin approximation $u_h \in V_h$ is determined as the solution of the discrete equation

$$A(u_h,\varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{1.5}$$

Then, the value $J(u_h)$ is an approximation for the target quantity $J(u)$. In practice, the existence of a solution $u \in V$ of the variational equation (1.4) has to be guaranteed by separate arguments outside the present general frame. Here, we only assume that the bilinear form $A(\cdot,\cdot)$ is sufficiently

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

regular on V and V_h, such that this solution as well as its approximation $u_h \in V_h$ is uniquely determined.

Our approach to estimating the functional error $J(u)-J(u_h)$ is based on embedding the given problem into the framework of optimal control. To this end, we consider the following trivial constraint optimization problem for $u \in V$:

$$J(u) = \min!, \quad A(u,\varphi) = F(\varphi), \quad \text{for all } \varphi \in V, \tag{1.6}$$

which is equivalent to evaluating $J(u)$ from the solution of $Au = f$. Introducing the corresponding Lagrangian $L(u,z) := J(u)+F(z)-A(u,z)$, with the 'adjoint' variable $z \in V$, the minimal solution u is characterized as the first component of a stationary point of $L(u,z)$. That is determined by the Euler–Lagrange system consisting of (1.4) and the adjoint problem

$$A(\varphi,z) = J(\varphi), \quad \text{for all } \varphi \in V. \tag{1.7}$$

Again, the actual existence of the adjoint solution $z \in V$ satisfying (1.7) is guaranteed in concrete situations by separate arguments. By construction, the solutions u and z are mutually adjoint to each other in the sense that $J(u) = A(u,z) = F(z)$. Hence, for computing $J(u)$, one could equally well try to compute $F(z)$. The Euler–Lagrange system is approximated by the Galerkin method in V_h resulting in the discrete equations (1.5) for $u_h \in V_h$ and

$$A(\varphi_h, z_h) = J(\varphi_h), \quad \text{for all } \varphi_h \in V_h,$$

for $z_h \in V_h$. For both errors $e := x-x_h$ and $e^* := z-z_h$, we have the Galerkin orthogonality property $A(e,\cdot) = 0 = A(\cdot,e^*)$ on V_h. Therefore, the mutual duality of u and z carries over to the corresponding errors,

$$J(e) = A(e,z) = A(e,e^*) = A(u,e^*) = F(e^*).$$

We introduce the notation $\rho(u_h,\cdot) := F(\cdot)-A(u_h,\cdot)$ and $\rho^*(z_h,\cdot) := J(\cdot)-A(\cdot,z_h)$ for the residuals of u_h and z_h, respectively. Then, again using Galerkin orthogonality, we have $\rho(u_h, z-\varphi_h) = A(e,e^*) = \rho^*(z_h, u-\varphi_h)$, with arbitrary $\varphi_h \in V_h$. From this, we see that

$$J(e) = \min_{\varphi_h \in V_h} \rho(u_h, z-\varphi_h) = \min_{\varphi_h \in V_h} \rho^*(z_h, u-\varphi_h) = F(e^*). \tag{1.8}$$

These error representations are to be evaluated computationally and serve as a basis for adaptive control of the discretization in computing $J(u)$. This duality of the primal error e and the adjoint error e^* is characteristic of the Galerkin method and is also found in nonlinear problems. The method for *a posteriori* error estimation described so far will be applied below in several concrete situations of 'conforming' Galerkin finite element discretizations of elliptic problems, also including an optimal control problem for which (1.6) becomes nontrivial.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

However, in many applications the Galerkin approximation of the given problem $Lu = f$ is based on modified variational formulations which are mesh-dependent. This is the case, for example, in the 'discontinuous' Galerkin method for transport-oriented or time-dependent problems and when standard Galerkin schemes are stabilized by introducing least-squares or streamline diffusion terms. In order to cover discretizations of this type too, the above framework has to be properly extended. In this context, the DWR method is mainly used as a formal guideline for deriving useful *a posteriori* error estimators. Because of the complexity of the concrete setting, this derivation may occasionally lack full mathematical rigour but eventually finds its justification by computational success.

The contents of this article are as follows. The concept underlying the DWR method will be introduced in Section 2 within an abstract setting. This serves as a basis for the application of the method to various rather different situations of variational problems. This will be made concrete in Section 3 for some linear model problems. In the same context, in Section 4 we will discuss some other approaches to functional-oriented error estimation which are based on 'energy norm' error estimates. The details of the practical realization of the DWR method will be addressed in Section 5. Though the derivation of the abstract *a posteriori* error estimates is equally simple for linear and nonlinear problems, in the latter case its practical evaluation poses particular difficulties. These will be discussed in Section 6. Section 7 deals with the application to Galerkin methods for time-dependent problems, particularly the heat equation and the acoustic wave equation.

The remaining sections are devoted to the application of the DWR method to various practical problems. In Section 8, we show an application to viscous incompressible fluid flow. Here, the main aspect is the interaction of different physical mechanisms with numerical stabilization, and its effect on the accuracy in computing a local quantity, in this case the 'drag coefficient'. The complexity of this physical model will be further increased in Section 9 by including compressibility effects due to temperature changes and chemical reactions. Section 10 contains an example from elasto-plasticity which involves a quasi-linear operator with nonlinearity not everywhere differentiable. This shows that the restrictive smoothness assumptions made in Section 2 are not really necessary for the method to work. Next, Section 11 describes the application of the DWR method to a nonstandard integro-differential equation governing the transfer of energy by radiation and scattering. This is an example in which, because of the high dimensionality of the problem, adaptive mesh design is indispensable. In Section 12, we present an application to a simple optimal control problem with linear state equation and quadratic cost functional. In this example the abstract framework of the DWR method introduced in Section 2 shows its full power.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Finally, Section 13 concludes the paper by presenting open problems and an outlook to future developments.

2. A paradigm for *a posterori* error control

We will develop the DWR method initially within an abstract functional analytic setting. The starting point is a general (though almost trivial) result on *a posteriori* error estimation in the Galerkin approximation of variational problems. From this, we then derive more detailed error representations for variational equations which yield residual-based error estimators with respect to functionals of the solution.

2.1. Galerkin approximation of stationary points

We begin with a general result for *a posteriori* error estimation in the Galerkin approximation of variational problems. The key point is that, here, the error is naturally measured in terms of the generating 'energy functional'. Let X be a function space and $L(\cdot)$ a differentiable functional on X. Its derivatives are denoted by $L'(\cdot;\cdot)$, $L''(\cdot;\cdot,\cdot)$ and $L'''(\cdot;\cdot,\cdot,\cdot)$. Here, we use the convention that in semilinear forms like $L'(\cdot;\cdot)$ the form is linear with respect to all arguments on the right of the semicolon. We seek a stationary point x of $L(\cdot)$ on X, that is,

$$L'(x;y) = 0, \quad \text{for all } y \in X. \tag{2.1}$$

This equation is approximated by a Galerkin method using a finite-dimensional subspace $X_h \subset X$, where $h \in \mathbb{R}_+$ is a discretization parameter. The discrete problem seeks $x_h \in X_h$ satisfying

$$L'(x_h;y_h) = 0, \quad \text{for all } y_h \in X_h. \tag{2.2}$$

To estimate the error $e := x - x_h$, we write

$$L(x) - L(x_h) \;=\; \tfrac{1}{2}L'(x_h;e) + \tfrac{1}{2}L'(x;e) + \tfrac{1}{2}\int_0^1 L''(x_h+se;e,e)\,\mathrm{d}s. \tag{2.3}$$

The second derivative $L''(x;\cdot,\cdot)$ is symmetric and, for convex $L(\cdot)$, we have $L''(x_h+se;e,e) \geq 0$. From this, we obtain

$$0 \;\leq\; L(x_h) - L(x) \;\leq\; -\tfrac{1}{2}L'(x_h;e), \tag{2.4}$$

and, using definition (2.2) of x_h,

$$0 \;\leq\; L(x_h) - L(x) \;\leq\; -\tfrac{1}{2}\min_{y_h \in X_h} L'(x_h;x-y_h). \tag{2.5}$$

The term $L'(x_h;x-y_h)$ represents a residual involving the weight factor $\omega := x - y_h$. Below, we will derive explicit representations for such terms depending on the concrete form of the underlying problem. This gives an

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

a posteriori bound for the error measured in terms of the energy functional $L(\cdot)$. Now, we consider the general case when the functional $L(\cdot)$ is not necessarily convex.

Proposition 2.1. For the Galerkin approximation of the variational problem (2.1), we have the *a posteriori* error representation

$$L(x) - L(x_h) \;=\; \tfrac{1}{2} \min_{y_h \in X_h} L'(x_h; x-y_h) + R. \tag{2.6}$$

The remainder term is given by

$$R := \tfrac{1}{2} \int_0^1 L'''(x_h+se; e, e, e)\, s(s-1)\, \mathrm{d}s$$

and vanishes if the functional $L(\cdot)$ is quadratic.

Proof. We introduce the notation

$$L'(\overline{xx_h}; y) := \int_0^1 L'(x_h+se; y)\, \mathrm{d}s.$$

and we begin with the trivial identity

$$L(x) - L(x_h) \;=\; L'(\overline{xx_h}; e) + \tfrac{1}{2}L'(x_h; e) - \tfrac{1}{2}L'(x_h; e) - \tfrac{1}{2}L'(x; e),$$

where we use the fact that $L'(x; e) = 0$. From this we obtain, by definition of x_h,

$$L(x) - L(x_h) \;=\; \tfrac{1}{2}L'(x_h; x-y_h) + L'(\overline{xx_h}; e) - \tfrac{1}{2}L'(x_h; e) - \tfrac{1}{2}L'(x; e),$$

with an arbitrary $y_h \in X_h$. The last two terms on the right are just the approximation of the second one by the trapezoidal rule. Recalling the corresponding remainder term

$$R \;=\; \tfrac{1}{2} \int_0^1 L'''(x_h+se; e, e, e)\, s(s-1)\, \mathrm{d}s,$$

we obtain the desired error representation (2.6). □

2.2. Galerkin approximation of variational equations

Now, we present an approach to *a posteriori* error estimation for the standard Galerkin approximation of variational equations based on the general result of Proposition 2.1. Let $A(\cdot;\cdot)$ be a differentiable semilinear form and $F(\cdot)$ a linear functional defined on some function space V. We seek a solution $u \in V$ to the variational equation

$$A(u; \varphi) = F(\varphi), \quad \text{for all } \varphi \in V. \tag{2.7}$$

For a finite-dimensional subspace $V_h \subset V$, again parametrized by $h \in \mathbb{R}_+$, the corresponding Galerkin approximation $u_h \in V_h$ is determined by

$$A(u_h; \varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{2.8}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

We assume that equations (2.7) and (2.8) possess (locally) unique solutions. Then, we have the Galerkin orthogonality relation

$$A(u;\varphi_h) - A(u_h;\varphi_h) = 0, \quad \varphi_h \in V_h. \tag{2.9}$$

The conventional approach to *a posteriori* error estimation in this situation is based on assumed coercivity properties of $A(\cdot;\cdot)$, namely

$$\|u - v\|_E \le c_s \sup_{z\in V,\,\|z\|=1} |A(u;z) - A(v;z)|, \quad u, v \in V, \tag{2.10}$$

with some generic 'energy norm' $\|\cdot\|_E$ on V. Then, using Galerkin orthogonality, we obtain a first *a posteriori* estimate for the error $e = u - u_h$:

$$\|e\|_E \le c_s \sup_{z\in V,\,\|z\|=1} \left\{ \min_{\varphi_h\in V_h} |\rho(u_h, z-\varphi_h)| \right\}, \tag{2.11}$$

where the residual $\rho(u_h;\cdot)$ is defined by

$$\rho(u_h;\varphi) := F(\varphi) - A(u_h;\varphi) \quad \varphi \in V. \tag{2.12}$$

We note that, by construction, $\rho(u_h;\cdot)$ vanishes on V_h.

Now, we consider again the case that *a priori* known coercivity properties of the form $A(\cdot;\cdot)$ are not available. Furthermore, we consider the more general situation of computing an approximation to $J(u)$, with a given differentiable functional $J(\cdot)$. We want to embed this situation into the general setting of Proposition 2.1. To this end, we note that the task of computing $J(u)$ from the solution of (2.7) can be equivalently formulated as solving a (trivial) constrained optimization problem for $u \in V$:

$$J(u) = \min!, \quad A(u;\varphi) = F(\varphi), \quad \text{for all } \varphi \in V. \tag{2.13}$$

Minima u correspond to stationary points $\{u, z\} \in V\times V$ of the Lagrangian

$$L(u;z) := J(u) + F(z) - A(u;z), \tag{2.14}$$

with the adjoint variable $z \in V$. Hence, we seek solutions $\{u, z\} \in V \times V$ to the Euler–Lagrange system

$$\begin{aligned} A(u;\varphi) &= F(\varphi), \quad \text{for all } \varphi \in V, \\ A'(u;\varphi, z) &= J'(u;\varphi), \quad \text{for all } \varphi \in V. \end{aligned} \tag{2.15}$$

Observe that the first equation of this system is just the given variational equation (2.7). In order to obtain a discretization of (2.15), we solve, in addition to (2.8), the following discrete adjoint equation:

$$A'(u_h;\varphi_h, z_h) = J'(u_h;\varphi_h), \quad \varphi_h \in V_h. \tag{2.16}$$

Again, we suppose that equations (2.15) and (2.16) possess unique solutions. To the solution $z_h \in V_h$ we associate the 'adjoint error' e^* and the

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

'adjoint residual'

$$\rho^*(z_h;\varphi) := J'(u_h;\varphi) - A'(u_h;\varphi,z_h), \quad \varphi \in V. \tag{2.17}$$

Proposition 2.2. For the Galerkin approximation of the Euler–Lagrange system (2.15), we have the *a posteriori* error representation

$$J(u) - J(u_h) \;=\; \tfrac{1}{2}\min_{\varphi_h\in V_h} \rho(u_h;z-\varphi_h) + \tfrac{1}{2}\min_{\varphi_h\in V_h} \rho^*(z_h;u-\varphi_h) + R \tag{2.18}$$

with the residual $\rho(u_h;\cdot)$ defined in (2.12) and the adjoint residual $\rho^*(z_h;\cdot)$ defined in (2.17). The remainder term is given by

$$\begin{aligned} R := \tfrac{1}{2}\int_0^1 \big\{ & J'''(u_h+se;e,e,e) - A'''(u_h+se;e,e,e,z_h+se^*) \\ & - 3A''(u_h+se;e,e,e^*)\big\}\, s(s-1)\,\mathrm{d}s \end{aligned} \tag{2.19}$$

and vanishes if $A(\cdot;\cdot)$ is linear and if $J(\cdot)$ is quadratic.

Proof. At the solutions $\{u,z\} \in V\times V$ and $\{u_h,z_h\} \in V_h\times V_h$,

$$L(u;z) - L(u_h;z_h) = J(u) - J(u_h).$$

Hence, Proposition 2.1, applied to the Lagrangian $L(\cdot;\cdot)$ with $X := V\times V$, $x = \{u,z\}$, $x_h = \{u_h,z_h\}$, *etc.*, yields a representation for the error $J(u) - J(u_h)$ in terms of the residuals $\rho(u_h;\cdot)$ and $\rho^*(z_h;\cdot)$. The remainder term has the form

$$\tfrac{1}{2}\int_0^1 L'''(x_h+se;\cdot,\cdot,\cdot)\, s(s-1)\,\mathrm{d}s.$$

Notice that $L(u;z)$ is linear in z. Consequently, the third derivative of $L(\cdot;\cdot)$ consists of only three terms, namely,

$$J'''(u_h+se;e,e,e) - A'''(u_h+se;e,e,e,z_h+se^*) - 3A''(u_h+se;e,e,e^*).$$

This completes the proof. □

The remainder term R in (2.19) is cubic in the errors e and e^* and may usually be neglected. The evaluation of the resulting error estimator

$$\eta_\omega(u_h,z_h) := \tfrac{1}{2}\min_{\varphi_h\in V_h} \rho(u_h;z-\varphi_h) + \tfrac{1}{2}\min_{\varphi_h\in V_h} \rho^*(z_h;u-\varphi_h)$$

requires approximations to the exact primal and dual solutions u and z. The error representation (2.18) is the nonlinear analogue of the (trivial) representation in the linear case derived in the introduction, as will be seen later. The relation between the primal and dual residuals is given in the next proposition.

Proposition 2.3. With the notation of Proposition 2.2,

$$\min_{\varphi_h\in V_h} \rho^*(z_h,u-\varphi_h) = \min_{\varphi_h\in V_h} \rho(u_h,z-\varphi_h) + \Delta\rho, \tag{2.20}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

with

$$\Delta\rho = \int_0^1 \left\{A''(u_h+see; e, e, z_h+se^*) - J''(u_h+se; e, e)\right\} \, \mathrm{d}s. \tag{2.21}$$

As a consequence, we have the following simplified error representation:

$$J(u) - J(u_h) = \min_{\varphi_h \in V_h} \rho(u_h, z - \varphi_h) + R, \tag{2.22}$$

with a quadratic remainder

$$R = \int_0^1 \left\{A''(u_h+se; e, e, z) - J''(u_h+se; e, e)\right\} s \, \mathrm{d}s.$$

Proof. Let us define

$$g(s) := J'(u_h+se; u-\varphi_h) - A'(u_h+se; u-\varphi_h, z_h+se^*).$$

Then we have

$$g(1) = J'(u; u-\varphi_h) - A'(u; u-\varphi_h, z) = 0,$$

by the definition of z. Further,

$$g(0) = J'(u_h; u-\varphi_h) - A'(u_h; u-\varphi_h, z_h) = \rho^*(z_h, u-\varphi_h).$$

The derivative of g is given by

$$\begin{aligned} g'(s) \quad = \quad & J''(u_h+se; e, u-\varphi_h) - A''(u_h+se; e, u-\varphi_h, z_h+se^*) \\ & -A'(u_h+se; u-\varphi_h, e^*). \end{aligned}$$

Therefore, we find that

$$\begin{aligned} \rho^*(z_h, u-\varphi_h) \quad = \quad & \int_0^1 \left\{A''(u_h+se; e, u-\varphi_h, z_h+se^*)\right. \\ & \left. -J''(u_h+se; e, u-\varphi_h)\right\} \mathrm{d}s \\ & + \int_0^1 A'(u_h+se; u-\varphi_h, e^*) \, \mathrm{d}s. \end{aligned}$$

Notice that the last term is just the primal residual, where we can substitute e^* by $z-\varphi_h$ with an arbitrary $\varphi_h \in V_h$.

The simplified error representation (2.22) can be derived by comparison of the remainder terms in (2.19) and (2.21) and some tedious computation. However, we prefer to give a direct argument. Integration by parts yields

$$R = -\int_0^1 \left\{A'(u_h+se; e, z) - J'(u_h+se; e)\right\} \, \mathrm{d}s + A'(u; e, z) - J'(u; e),$$

and the last term vanishes by definition of z. Therefore, we have

$$R = J(u) - J(u_h) + \rho(u_h, z).$$

Noticing that $\rho(u_h, z) = \rho(u_h, z-\varphi_h)$, for $\varphi_h \in V_h$, completes the proof. □

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Remark 2.1. We note that the results of Propositions 2.2 and 2.3 hold true in the following more general situation. Let V and W be Banach spaces. We consider a Lagrangian $L : V \times W \to \mathbb{R}$ defined as in (2.14). The corresponding stationary point $\{u, z\}$ is approximated by a Galerkin method using subspaces $V_h \subset V$ and $W_h \subset W$. Then, for the corresponding discrete solution $\{u_h, z_h\}$, we have

$$J(u) - J(u_h) = \tfrac{1}{2} \min_{\varphi_h \in W_h} \rho(u_h; z - \varphi_h) + \tfrac{1}{2} \min_{\varphi_h \in V_h} \rho^*(z_h; u - \varphi_h) + R,$$

where R has the same form as in (2.19). This generalization allows us to include Petrov–Galerkin schemes into the general framework.

In the case of a *linear* functional $J(\cdot)$ and *linear* variational equation, Proposition 2.3 states that the two residuals coincide. In general, the difference is quadratic in e and can be supposed to be relatively small in many practical situation. In general, the difference $\Delta\rho$ can be used as an indicator of the influence of the nonlinearity on the error.

In the following proposition, we derive an alternative representation where no remainder occurs.

Proposition 2.4. For the Galerkin approximation (2.8) we have the *a posteriori* error representation

$$J(u) - J(u_h) = \min_{\varphi_h \in V_h} \rho(u_h; z - \varphi_h), \tag{2.23}$$

where z is defined by the adjoint problem

$$\int_0^1 A'(u_h + se; \varphi, z)\, \mathrm{d}s = \int_0^1 J'(u_h + se; \varphi)\, \mathrm{d}s, \quad \text{for all } \varphi \in V. \tag{2.24}$$

Proof. By elementary calculus,

$$A(u; \varphi) - A(u_h; \varphi) = \int_0^1 A'(u_h + se; e, \varphi)\, \mathrm{d}s = A'(\overline{uu_h}; e, \varphi),$$

and analogously for $J(u) - J(u_h)$. Taking $\varphi = e$ in the adjoint problem (2.24), we obtain

$$A'(\overline{uu_h}; e, z) = J'(\overline{uu_h}; e),$$

and combining this with the previous identities,

$$J(u) - J(u_h) = A'(\overline{uu_h}; e, z) = A(u; z) - A(u_h; z).$$

Now, using the Galerkin orthogonality (2.9) yields

$$J(u) - J(u_h) = A(u; z - \varphi_h) - A(u_h; z - \varphi_h) = F(z - \varphi_h) - A(u_h; z - \varphi_h),$$

with an arbitrary $\varphi_h \in V_h$. This implies (2.23). □

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Remark 2.2. We note that the result of Proposition 2.4 can also be inferred from the general Proposition 2.1. To this end, we construct an artificial optimization problem which is equivalent to the present situation. Let $\hat{u}$ and $\hat{u}_h$ be two fixed elements of V and set $\hat{e} := \hat{u} - \hat{u}_h$. The linear Lagrangian

$$\tilde{L}(u; z) := \int_0^1 J'(\hat{u}_h + s\hat{e})(\hat{u} - u))\,\mathrm{d}s + \int_0^1 A'(\hat{u}_h + s\hat{e}; \hat{u} - u, z)\,\mathrm{d}s$$

has stationary point $\{\tilde{u}, \tilde{z}\}$ defined by

$$A(\tilde{u}; \varphi) = A(\hat{u}; \varphi), \quad \text{for all } \varphi \in V,$$

and

$$A'(\overline{\hat{u}\hat{u}_h}; \varphi, \tilde{z}) = J'(\overline{\hat{u}\hat{u}_h}; \varphi), \quad \text{for all } \varphi \in V.$$

Let $\{\tilde{u}_h, \tilde{z}_h\} \in V_h \times V_h$ be the Galerkin approximation to these equations. If we now choose $\hat{u} = u$ and $\hat{u}_h = u_h$, we see that $\tilde{u} = u$ and $\tilde{z} = z$, and the two previous equations reduce to the continuous equations (2.7) and (2.24). Furthermore, their discrete analogues generate equation (2.8) for u_h and

$$A'(\overline{uu_h}; \varphi, z_h) = J'(\overline{uu_h}; \varphi_h), \quad \text{for all } \varphi_h \in V_h$$

for $z_h \in V_h$. Now applying the general result of Proposition 2.1, we find that

$$J(u) - J(u_h) = \tfrac{1}{2}\tilde{\rho}(u_h; z - \varphi_h) + \tfrac{1}{2}\tilde{\rho}^*(z_h; u - \varphi_h),$$

since the remainder vanishes because of the linearity of $\tilde{L}(\cdot;\cdot)$. By definition, we have that $\tilde{\rho}(u_h; \varphi) = \rho(u_h; \varphi)$ and $\tilde{\rho}^*(z_h; u) = J(u) - J(u_h)$, which completes the argument.

The result of Proposition 2.4 cannot be used directly in practice, since the underlying adjoint problem (2.24) involves the unknown solution u. However, in order to study the effect of the nonlinearity, one might be willing to approximate (2.24) by means of some approximation $\hat{u}_h$ obtained in a possibly richer space $\hat{V}_h \subset V$, that is,

$$A'(\overline{\hat{u}_h u_h}; \varphi, z) = J'(\overline{\hat{u}_h u_h}; \varphi), \quad \text{for all } \varphi \in V. \tag{2.25}$$

For the simplest choice $\hat{u}_h = u_h$, (2.25) reduces to

$$A'(u_h; \varphi, z) = J'(u_h; \varphi), \quad \text{for all } \varphi \in V, \tag{2.26}$$

and we are in the situation of Proposition 2.2. In the following, we provide a theoretical result on the effect of using the approximate adjoint problem (2.25).

Proposition 2.5. Suppose that $\hat{u}_h \in \hat{V}_h$ is an improved approximation to the solution u with corresponding error $\hat{e} := u - \hat{u}_h$. Let $z \in V$ be

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

the solution of the linearized adjoint problem (2.25). Then, for the Galerkin scheme (2.8), we have the error identity

$$J(u) - J(u_h) \;=\; \min_{\varphi_h \in V_h} \rho(u_h; z-\varphi_h) + R \tag{2.27}$$

with a remainder bounded by

$$|R| \;\le\; \tfrac{1}{2} \max_{\xi \in \overline{u\hat{u}_h u_h}} \big|J''(\xi;\hat{e},e) - A''(\xi;\hat{e},e,z)\big|,$$

where $\overline{u\hat{u}_h u_h}$ denotes the 'triangle' in V spanned by the three points $\{u, \hat{u}_h, u_h\}$. The remainder term vanishes if $A(\cdot;\cdot)$ and $J(\cdot)$ are linear.

Proof. We set $\hat{e}_h := \hat{u}_h - u_h$. Then, again by elementary calculus,

$$\begin{aligned} J(u) - J(u_h) \;&=\; J'(\overline{\hat{u}_h u_h}; e) + \int_0^1 \big\{J'(u_h + se; e) - J'(u_h + s\hat{e}_h; e)\big\}\,\mathrm{d}s, \\ A(u;z) - A(u_h;z) \;&=\; A'(\overline{\hat{u}_h u_h}; e, z) \\ &\qquad + \int_0^1 \big\{A'(u_h + se; e, z) - A'(u_h + s\hat{e}_h; e, z)\big\}\,\mathrm{d}s. \end{aligned}$$

Taking $\varphi = e$ in the adjoint problem (2.25), we conclude from the previous identities that

$$J(u) - J(u_h) = A(u;z) - A(u_h;z) + R(\overline{u\hat{u}_h u_h}; \hat{e}, e, z),$$

with the remainder term

$$\begin{aligned} R := &\int_0^1 \big\{J'(u_h + se; e) - J'(u_h + s\hat{e}_h; e)\big\}\,\mathrm{d}s \\ &- \int_0^1 \big\{A'(u_h + se; e, z) - A'(u_h + s\hat{e}_h; e, z)\big\}\,\mathrm{d}s. \end{aligned}$$

Observing that $u_h + se - u_h + s\hat{e}_h = s\hat{e}$, the first integral on the right is rewritten as

$$\begin{aligned} &\int_0^1 \big\{J'(u_h + se; e) - J'(u_h + s\hat{e}_h; e)\big\}\,\mathrm{d}s \\ &\qquad = -\int_0^1 \left\{ \int_0^1 J''(u_h + s\hat{e}_h + ts\hat{e}; s\hat{e}, e)\,\mathrm{d}t \right\} \mathrm{d}s, \end{aligned}$$

and analogously for the second one. We note that $u_h + s\hat{e}_h + ts\hat{e} = (1-s)u_h + tsu + s(1-t)\hat{u}_h$ is a convex combination of the three points. Hence,

$$|R| \;\le\; \tfrac{1}{2} \max_{\xi \in \overline{u\hat{u}_h u_h}} \big|J''(\xi;\hat{e},e) - A''(\xi;\hat{e},e,z)\big|,$$

as asserted. Now, using the Galerkin orthogonality (2.9) again yields

$$J(u) - J(u_h) = \rho(u_h; z-\varphi_h) + R,$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

with an arbitrary $\varphi_h \in V_h$. From this, we conclude (2.27). Clearly, the remainder term vanishes if $A(\cdot;\cdot)$ and $J(\cdot)$ are linear. □

Assuming that $\hat{e}$ is much smaller than e, the remainder term $R(\overline{uu_h}; \hat{e}, e, z)$ may be neglected in comparison with the residual term

$$\hat{\eta}_\omega(u_h) := \min_{\varphi_h \in V_h} \rho(u_h; z - \varphi_h).$$

This result provides the basis for successive improvement of the reliability of the error estimator $\hat{\eta}_\omega(u_h)$ by using improved approximations to $\hat{u}_h$ in the linearization. This concept will be discussed in more detail below for a concrete situation.

2.3. Extension to nonstandard Galerkin methods

In many concrete situations the Galerkin approximation is based on a modified variational formulation which is mesh-dependent. This is the case, for example, in the 'discontinuous' Galerkin method for transport-oriented or time-dependent problems (see Section 7) and also when a standard Galerkin scheme is stabilized by introducing least-squares or streamline diffusion terms (see Sections 3, 8, 9, and 11). In order to cover discretizations of this type too, the framework developed so far has to be properly extended.

Let $\hat{V}$ be some larger space containing the solution space V, and, in particular, the solution $u \in V$, of the problem

$$A(u;\varphi) = F(\varphi), \quad \text{for all } \varphi \in V, \tag{2.28}$$

as well as the finite-dimensional space V_h. In most concrete situations, we simply set $\hat{V} := V \oplus V_h$. On $\hat{V}$, we use an h-dependent 'stabilization form' $S_h(\cdot;\cdot)$ in defining

$$A_h(\cdot;\cdot) := A(\cdot;\cdot) + S_h(\cdot;\cdot),$$

and consider the approximate problem

$$A_h(u_h;\varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h \in V_h, \tag{2.29}$$

We assume again that the modified form $A_h(\cdot;\cdot)$ is regular on V_h, such that the finite-dimensional problem (2.29) has a unique solution $u_h \in V_h$. The residual of this solution $u_h \in V_h$ is defined by $\rho_h(u_h,\cdot) := F(\cdot) - A_h(u_h;\cdot)$. Further, this discretization is required to be 'consistent' with the original problem in the sense that the solution $u \in V$ of (2.28) automatically satisfies (2.29),

$$A_h(u;\varphi) = F(\varphi), \quad \varphi \in V.$$

This immediately implies Galerkin orthogonality in the form

$$\rho_h(u_h;\varphi_h) = A_h(u;\varphi_h) - A_h(u_h;\varphi_h) = 0, \quad \varphi_h \in V_h. \tag{2.30}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

This consistency condition may have to be relaxed in some concrete situations introducing additional perturbation terms in the resulting *a posteriori* error estimates. For controlling the error $J(u)-J(u_h)$, we consider the h-dependent adjoint problem

$$\tilde{A}_h'(u;\varphi,z) := A'(u;\varphi,z) + \tilde{S}_h'(u;\varphi,z) \;=\; J'(u;\varphi), \quad \text{for all } \varphi \in \hat{V}, \quad (2.31)$$

where $\tilde{S}_h'(u;\cdot,\cdot)$ may consist of only a definite part of $S_h'(u;\cdot,\cdot)$ chosen such that (2.31) can be guaranteed to possess a unique solution $z \in \hat{V}$. Clearly, by its construction this adjoint solution z usually also depends on h.

Remark 2.3. The previous construction requires some comment. For example, in the case of a discontinuous Galerkin scheme the adjoint solution z usually exists in the original solution space V and is independent of h. On the other hand, in the presence of least-squares stabilization the adjoint solution may exist only in a weak sense in $\hat{V}$ and its limit behaviour for $h \to 0$ is not clear. However, this may not be too critical, since in the practical evaluation of the residual $\rho_h(u_h; z-\varphi_h)$ the weights $z-\varphi_h$ are, after all, approximated using the solution $z_h \in V_h$ of the discretized adjoint problem

$$\tilde{A}_h'(u_h;\varphi_h,z_h) = J'(u_h;\varphi_h), \quad \text{for all } \varphi_h \in V_h. \quad (2.32)$$

This problem, in turn, can often be interpreted as stabilized approximation of a formal adjoint problem in terms of the given differential operator.

In the following, we derive an analogue of the error representation in Proposition 2.3 for the present situation.

Proposition 2.6. With the above notation, we have the error representation

$$J(u) - J(u_h) \;=\; \min_{\varphi_h \in V_h} \rho_h(u_h, z-\varphi_h) + R_h, \quad (2.33)$$

with the remainder term

$$R_h = (\tilde{S}_h'-S_h')(u;e,z) + \int_0^1 \{A_h''(u_h+se;e,e,z) - J''(u_h+se;e,e)\} s \,\mathrm{d}s.$$

Proof. We note that

$$J(u) - J(u_h) \;=\; J'(u;e) - \int_0^1 J''(u_h+se;e,e)s \,\mathrm{d}s.$$

Taking $\varphi := e$ in (2.31) and using the result in the previous equation gives us

$$\begin{aligned} J(u) - J(u_h) \;&=\; \tilde{A}_h'(u;e,z) - \int_0^1 J''(u_h+se;e,e)s \,\mathrm{d}s \\ &=\; A_h'(u;e,z) + (\tilde{S}_h'-S_h')(u;e,z) - \int_0^1 J''(u_h+se;e,e)s \,\mathrm{d}s. \end{aligned}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

We combine this with the relation

$$\begin{aligned} A_h'(u;e,z) &= \int_0^1 A_h'(u_h+se;e,z)\,\mathrm{d}s + \int_0^1 A_h''(u_h+se;e,e,z)s\,\mathrm{d}s \\ &= \rho_h(u_h;z) + \int_0^1 A_h''(u_h+se;e,e,z)s\,\mathrm{d}s, \end{aligned}$$

and use the Galerkin orthogonality (2.30) to obtain

$$\begin{aligned} J(u)-J(u_h) &= \rho_h(u_h;z-\varphi_h) + (\tilde{S}_h'-S_h')(u;e,z) \\ &\quad + \int_0^1 \{A_h''(u_h+se;e,e,z) - J''(u_h+se;e,e)\}s\,\mathrm{d}s. \end{aligned}$$

with an arbitrary $\varphi_h \in V_h$. This completes the proof. □

In concrete situations the first part of the remainder term, $(\tilde{S}_h'-S_h')(u;e,z)$, contains small parameters and can be neglected in comparison with the residual term $\rho_h(u_h;z-\varphi_h)$.

Notes and references

Global energy and L^2-norm error estimates for nonlinear problems exploiting coercivity properties or assuming bounds for Frechét derivatives have been derived, for instance, by Caloz and Rappaz (1997), Medina, Picasso and Rappaz (1996), and Verfürth (1993, 1994, 2000). Duality arguments as described above are very common in the *a priori* error analysis of Galerkin finite element methods (see, *e.g.*, Ciarlet (1978) and Brenner and Scott (1994)). Their use for *a posteriori* error estimation particularly in nonlinear problems was suggested by Johnson (1994), Eriksson and Johnson (1995*b*), and Johnson and Rannacher (1994) (see also Eriksson *et al.* (1995)). The presentation based on optimal control principles uses ideas from Rannacher (200x) and Becker, Kapp and Rannacher (2000*b*).

3. The dual-weighted residual method

Having prepared the general frame for *a posteriori* error estimation in Section 2, we now discuss the application of these concepts to concrete situations. For illustration, we first consider some simple model situations: the Poisson equation, an eigenvalue problem of the Laplace operator and a linear transport equation. This setting is also used to introduce the basic facts about Galerkin finite element approximation, as they will be used throughout the paper. Below, we will need some notation from the theory of function spaces. Readers who are familiar with the usual Lebesgue- and Sobolev-space notation may want to skip this and continue with the next subsection on finite element approximation.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Notation for function spaces
For a domain $Q \subset \mathbb{R}^d$, we let $L^2(Q)$ denote the Lebesgue space of square-integrable functions on Q, which is a Hilbert space with the scalar product and norm

$$(v,w)_Q = \int_Q v\,w\,\mathrm{d}x \qquad \text{and} \qquad \|v\|_Q = \Big(\int_Q |v|^2\,\mathrm{d}x\Big)^{1/2}.$$

Analogously, $L^2(\partial Q)$ denotes the space of square-integrable functions defined on the boundary ∂Q equipped with the scalar product and norm

$$(v,w)_{\partial Q} = \int_{\partial Q} v\,w\,\mathrm{d}s \qquad \text{and} \qquad \|v\|_{\partial Q} = \Big(\int_{\partial Q} |v|^2\,\mathrm{d}s\Big)^{1/2}.$$

The Sobolev spaces $H^1(Q)$ and $H^2(Q)$ consist of the functions $v \in L^2(Q)$ that possess first- and second-order (distributional) derivatives $\nabla v \in L^2(Q)^d$ and $\nabla^2 v \in L^2(Q)^{d\times d}$, respectively. For functions in these spaces, we use the seminorms

$$\|\nabla v\|_Q = \Big(\int_Q |\nabla v|^2\,\mathrm{d}x\Big)^{1/2}, \qquad \|\nabla^2 v\|_Q = \Big(\int_Q |\nabla^2 v|^2\,\mathrm{d}x\Big)^{1/2}.$$

The space $H^1(Q)$ can be embedded in the space $L^2(\partial Q)$, such that for each $v \in H^1(Q)$ there exists a trace $v_{|\partial Q} \in L^2(\partial Q)$. Further, the functions in the subspace $H^1_0(Q) \subset H^1(Q)$ are characterized by the property $v_{|\partial Q} = 0$. By the Poincaré inequality,

$$\|v\|_Q \le c\,\|\nabla v\|_Q, \quad v \in H^1_0(Q), \tag{3.1}$$

the H^1-seminorm $\|\nabla v\|_Q$ is a norm on the subspace $H^1_0(Q)$. If the set Q is the set Ω on which the differential equation is posed, we usually omit the subscript Ω in the notation 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 $v : \Omega \to \mathbb{R}^d$ or $\mathbb{R}^{d\times d}$.

3.1. Finite element discretization of the Poisson equation

We begin with the model problem

$$-\Delta u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \tag{3.2}$$

posed on a polygonal domain $\Omega \subset \mathbb{R}^2$. The natural solution space for this boundary value problem is the Sobolev space $V = H^1_0(\Omega)$. The variational formulation of (3.2) seeks $u \in V$ such that

$$A(u,\varphi) = (f,\varphi), \quad \text{for all } \varphi \in V, \tag{3.3}$$

where $A(u,\varphi) := (\nabla u, \nabla\varphi)$. The finite element approximation of (3.3) uses

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

finite-dimensional subspaces

$$V_h = \{v \in V : v_{|K} \in P(K),\ K \in \mathbb{T}_h\},$$

defined on decompositions $\mathbb{T}_h$ of Ω into triangles or quadrilaterals K (*cells*) of width $h_K = \mathrm{diam}(K)$; we write $h = \max_{K\in\mathbb{T}_h} h_K$ for the *global* mesh width. Here, $P(K)$ denotes a suitable space of polynomial-like functions defined on the cell $K \in \mathbb{T}_h$. In the numerical results discussed below, we have mostly used 'bilinear' finite elements on quadrilateral meshes in which case $P(K) = \tilde{Q}_1(K)$ consists of shape functions obtained via a bilinear transformation from the space of bilinear functions $Q_1(\hat{K}) = \mathrm{span}\{1, x_1, x_2, x_1x_2\}$ on the reference cell $\hat{K} = [0,1]^2$. Local mesh refinement or coarsening is realized by using *hanging nodes*, in such a way that global conformity is preserved, that is, $V_h \subset V$. For technical details of finite element spaces, the reader may consult the standard literature, for instance Ciarlet (1978) or Brenner and Scott (1994), and, particularly for the treatment of hanging nodes, Carey and Oden (1984). Now, approximations $u_h \in V_h$ are determined by

$$A(u_h, \varphi_h) = (f, \varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{3.4}$$

The corresponding residual and error of $u_h \in V_h$ are $\rho(u_h, \cdot) := (f, \cdot) - A(u_h, \cdot)$ and $e = u - u_h$, respectively. By construction, we have 'Galerkin orthogonality'

$$\rho(u_h, \varphi_h) = A(e, \varphi_h) = 0, \quad \varphi_h \in V_h. \tag{3.5}$$

In order to convert (3.4) into an algebraic equation, one uses a 'nodal basis' $\{\varphi_h^i\}_{i=1,\dots,n}$ ($n := \dim V_h$) of the finite element space V_h. The coefficient vector $x_h = (x_i)_{i=1}^n$ in the expansion $u_h = \sum_{i=1}^n x^i \varphi_h^i$ is determined by the linear system

$$A_h\, x_h = b_h, \tag{3.6}$$

with the 'stiffness matrix' $A = (a_{ij})_{i,j=1}^n$ and the 'load vector' $b_h = (b_i)_{i=1}^n$ defined by $a_{ij} = A(\varphi_h^j, \varphi_h^i)$ and $b_i = (f, \varphi_h^i)$.

3.1.1. A priori error analysis

We begin with a brief review of the *a priori* error analysis for the scheme (3.4). We let $i_h u \in V_h$ denote the natural 'nodal interpolation' of $u \in C(\bar{\Omega})$ satisfying $i_h u(P) = u(P)$ at all nodal points P. We have

$$\|\nabla(u - i_h u)\|_K \ \le\ c_i h_K \|\nabla^2 u\|_K \tag{3.7}$$

(see, *e.g.*, Ciarlet (1978) and Brenner and Scott (1994)) and, furthermore,

$$\|u - i_h u\|_K + h_K^{1/2} \|u - i_h u\|_{\partial K} \ \le\ c_i h_K^2 \|\nabla^2 u\|_K, \tag{3.8}$$

for certain positive 'interpolation constants' c_i.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

(i) By the projection property of the Galerkin finite element scheme the interpolation estimate (3.7) directly implies the 'energy error estimate'

$$\|\nabla e\| \;=\; \inf_{\varphi_h\in V_h} \|\nabla(u-\varphi_h)\| \;\le\; c_i h^2 \|\nabla^2 u\|. \tag{3.9}$$

(ii) Further, employing a duality argument ('Aubin–Nitsche trick'),

$$-\Delta z = \|e\|^{-1} e \quad \text{in } \Omega, \quad z = 0 \;\text{ on } \partial\Omega, \tag{3.10}$$

we obtain

$$\|e\| = (e,-\Delta z) = A(e,z) = A(e,z-i_h z) \le c_i c_s h \|\nabla e\|, \tag{3.11}$$

where the constant c_s is defined by the *a priori* bound $\|\nabla^2 z\| \le c_s$. In view of the energy error estimate (3.9), this implies the 'L^2-error estimate'

$$\|e\| \;\le\; c_i^2 c_s h^2 \|\nabla^2 u\|. \tag{3.12}$$

(iii) On the basis of the global error estimates (3.9) and (3.12), we can also obtain error bounds for various other quantities, *e.g.*, point values or boundary moments (in $\mathbb{R}^2$):

$$|\nabla e(P)| \approx h^{-1}\|\nabla e\|, \qquad |(e,\psi)_{\partial\Omega}| \approx h^{-1/2}\|e\|.$$

All these estimates are only suboptimal in h. This tells us that estimating the error in functional output usually requires special effort and cannot simply be reduced to the standard energy or L^2-error estimates.

3.1.2. A posteriori error analysis

Next, we derive *a posteriori* error estimates. Suppose that the error is to be controlled with respect to some (linear) 'error functional' $J(\cdot)$ defined on V. Let $z \in V$ be the solution of the corresponding adjoint problem

$$A(\varphi, z) = J(\varphi), \quad \text{for all } \varphi \in V, \tag{3.13}$$

and $z_h \in V_h$ its finite element approximation defined by

$$A(\varphi_h, z_h) = J(\varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{3.14}$$

The 'adjoint' error is denoted by $e^* := z - z_h$. Applying the abstract *a posteriori* error representation of Proposition 2.3 to this situation yields

$$J(e) = \rho(u_h, z-\varphi_h), \tag{3.15}$$

for arbitrary $\varphi_h \in V_h$. Now, the residual expression for u_h has to be further developed. By cell-wise integration by parts, we obtain

$$\rho(u_h, z-\varphi_h) \;=\; \sum_{K\in\mathbb{T}_h} \big\{(f+\Delta u_h, z-\varphi_h)_K - (n\cdot\nabla u_h, z-\varphi_h)_{\partial K}\big\}. \tag{3.16}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

At this point, we have assumed that the domain Ω is polygonal (or polyhedral) in order to ease the approximation of the boundary $\partial\Omega$. In the presence of curved parts of $\partial\Omega$ the formula (3.16) contains additional terms representing the error caused by the polygonal approximation of the boundary (see, *e.g.*, Becker and Rannacher (1996*b*)). In the tests below, these terms are suppressed since they are usually negligibly small. The representation (3.16) can be rewritten by combining the contributions from cell edges in the following form:

$$J(e) \;=\; \sum_{K\in\mathbb{T}_h} \big\{(R(u_h), z-\varphi_h)_K - (r(u_h), z-\varphi_h)_{\partial K}\big\}, \tag{3.17}$$

with the cell and edge residuals $R(u_h)$ and $r(u_h)$, respectively, defined by

$$R(u_h)_{|K} = f + \Delta u_h, \qquad r(u_h)_{|\Gamma} := \begin{cases} \frac{1}{2} n\cdot[\nabla u_h], & \text{if } \Gamma \subset \partial K \setminus \partial\Omega, \\ 0, & \text{if } \Gamma \subset \partial\Omega, \end{cases}$$

where $[\nabla u_h]$ denotes the jump of ∇u_h across the inter-element edges Γ, and $z - \varphi_{h|\Gamma} = 0$ on $\Gamma \subset \partial\Omega$ is observed. From the error identity (3.17), we can infer the following result.

Proposition 3.1. For the finite element approximation of the Poisson equation (3.2), we have the *a posteriori* error estimate

$$|J(e)| \;\le\; \eta_\omega(u_h) := \sum_{K\in\mathbb{T}_h} \rho_K\, \omega_K, \tag{3.18}$$

where the cell residuals ρ_K and weights ω_K are given by

$$\rho_K := \|R(u_h)\|_K + h_K^{-1/2}\|r(u_h)\|_{\partial K},$$
$$\omega_K := \|z-\varphi_h\|_K + h_K^{1/2}\|z-\varphi_h\|_{\partial K},$$

with a suitable approximation $\varphi_h \in V_h$ to z.

Remark 3.1. We give an interpretation of the *a posteriori* error estimate (3.18). On each cell, we have an 'equation residual' $R(u_h) := f + \Delta u_h$ and a 'flux residual' $r(u_h) := n\cdot[\nabla u_h]$, the latter one expressing smoothness of the discrete solution. Both residuals can easily be evaluated. They are multiplied by the weighting function $z - \varphi_h$, which provides quantitative information about the impact of these cell residuals on the error $J(e)$ in the target quantity. In this sense $z-\varphi_h$ may be viewed as *sensitivity factors* as in optimal control problems. Setting $\varphi_h = i_h z$, we have

$$\frac{\partial J(e)}{\partial \rho_K} \;\approx\; \omega_K \;\approx\; c_i h_K^2 \|\nabla^2 z\|_K.$$

Accordingly, the influence factors have the behaviour $h_K^2\|\nabla^2 z\|_K$, characteristic of the Galerkin finite element approximation ('projection method').

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

This allows us to exploit the 'Galerkin orthogonality' in deriving the local weights ω_K. We note that in a standard finite difference or finite volume discretization of (3.2), lacking the full Galerkin orthogonality property, the corresponding *influence factors* would behave like $\omega_K \approx h_K \|\nabla z\|_K$. This reflects the lack of accuracy of these methods in computing the target quantity.

Remark 3.2. We have developed the error estimate (3.18) into a rather compact form in order to make its structure transparent. For practical use in mesh adaptation, we suggest avoiding taking norms but, rather, directly evaluating the error representation in the form

$$\eta_\omega(u_h) = \sum_{K\in\mathbb{T}_h} \eta_K = \sum_{K\in\mathbb{T}_h} \big|(R(u_h), z-\varphi_h)_K - (r(u_h), z-\varphi_h)_{\partial K}\big|. \quad (3.19)$$

Corresponding strategies will be discussed in Section 5.

Remark 3.3. The evaluation of the *a posteriori* error bound (3.18) requires us to provide sufficiently accurate approximations to the adjoint solution z. Techniques for generating these will be discussed in Section 5. In Proposition 3.1, we have only used the residual term corresponding to u_h. In view of the identities (3.15), we could also have used the relation

$$J(e) = \min_{\varphi_h\in V_h} \rho^*(z_h, u-\varphi_h) = (f, e^*),$$

resulting in the *a posteriori* error bound

$$|(f, e^*)| \;\le\; \eta^*_\omega(z_h) := \sum_{K\in\mathbb{T}_h} \rho^*_K\, \omega^*_K,$$

with the residual terms and weights, defined analogously,

$$\rho^*_K := \|R^*(z_h)\|_K + h_K^{-1/2}\|r^*(z_h)\|_{\partial K},$$
$$\omega^*_K := \|u-\varphi_h\|_K + h_K^{1/2}\|u-\varphi_h\|_{\partial K},$$

with a suitable approximation $\varphi_h \in V_h$ to u. Obviously, the evaluation of this error bound requires us to compute sufficiently accurate approximations to the solution u. Which approach to the estimation of $|J(e)|$ is used, the one based on u_h or the alternative one based on z_h, may depend on how expensive the construction of more accurate approximations to z or u will be. Roughly speaking, the underlying principle is as follows. Computing a more accurate approximation to z yields an accurate estimate for $J(e)$, but then an even better approximation to $J(u)$ should be obtained by evaluating (f, z). However, to ensure this improved approximation accuracy by an *a posteriori* error estimate, in turn, requires an even better approximation to u. Which way to proceed in practice depends on the different regularity properties of u and z. In the examples presented below, particularly the

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

nonlinear ones, we have chosen to base error estimation as well as mesh adaptation on an *a posteriori* error estimate in terms of the approximate solution u_h.

Remark 3.4. For the first-degree finite elements considered here, the contribution from the edge residual $r(u_h)$ dominates that from the equation residual $R(u_h)$ and the latter can therefore be neglected in most cases (see Carstensen and Verfürth (1999) and Becker and Rannacher (1996*b*)). Further, in the case of a curved boundary or nonzero boundary data, the error estimator $\eta_\omega(u_h)$ in (3.18) contains additional terms measuring the effects of the approximation along the boundary. In regular cases these terms are usually smaller than the edge terms and will be suppressed in the following for simplicity (for a rigorous *a posteriori* error analysis of boundary approximation see Dörfler and Rumpf (1998)).

Remark 3.5. The *a posteriori* error estimate (3.18) directly extends to more general diffusion operators with variable coefficients,

$$Lu := -\nabla\cdot\{a\nabla u\} + bu = f \quad \text{in } \Omega,$$

and mixed boundary conditions, $\partial\Omega = \Gamma_D \cup \Gamma_N$,

$$u_{|\Gamma_D} = 0, \quad n\cdot\{a\nabla u_h\}_{|\Gamma_N} = g.$$

The only difference is in the definition of the equation and edge residuals, which take the form $R(u_h)_{|K} = f + \nabla\cdot\{a\nabla u_h\} - bu_h$ and

$$r(u_h)_{|\Gamma} := \begin{cases} \frac{1}{2}n\cdot[a\nabla u_h], & \text{if } \Gamma \subset \partial K\setminus\partial\Omega, \\ 0, & \text{if } \Gamma \subset \Gamma_D, \\ n\cdot\{a\nabla u_h\} - g, & \text{if } \Gamma \subset \Gamma_N, \end{cases}$$

respectively. The incorporation of non-homogeneous Dirichlet boundary conditions will be discussed below in the context of concrete examples.

3.1.3. Examples

We want to illustrate some features of the DWR method by two simple examples.

Example 1: Computation of mean boundary stress. The first example is meant as an illustrative exercise. For problem (3.2) posed on the unit circle $\Omega := \{x \in \mathbb{R}^2 : |x| < 1\}$, we consider the functional

$$J(u) = \int_{\partial\Omega} \partial_n u \,\mathrm{d}s \quad \left(= -\int_\Omega f \,\mathrm{d}x \right),$$

and ask the question: What is an optimal mesh-size distribution for computing $J(u)$? The corresponding adjoint problem

$$a(\varphi, z) = (1, \partial_n\varphi)_{\partial\Omega}, \quad \text{for all } \varphi \in V \cap C^1(\bar\Omega)$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

has a measure solution with density of the form $z \equiv -1$ in Ω, $z = 0$ on $\partial\Omega$. In order to avoid the use of measures, we consider the regularized functional

$$J_\varepsilon(\varphi) = |S_\varepsilon|^{-1} \int_{S_\varepsilon} \partial_n \varphi \, \mathrm{d}x = \int_{\partial\Omega} \partial_n \varphi \, \mathrm{d}s + \mathcal{O}(\varepsilon), \quad \varepsilon = \mathrm{TOL},$$

where $S_\varepsilon = \{x \in \Omega : \mathrm{dist}\{x, \partial\Omega\} < \varepsilon\}$. Then, the adjoint solution is given by

$$z_\varepsilon = \begin{cases} -1, & \text{in } \Omega \setminus S_\varepsilon, \\ -\varepsilon^{-1}\mathrm{dist}\{x, \partial\Omega\}, & \text{in } S_\varepsilon. \end{cases}$$

This implies that

$$J_\varepsilon(e) \le c_i \sum_{K \in \mathbb{T}_h, K \cap S_\varepsilon \neq \emptyset} \rho_K \, h_K^2 \|\nabla^2 z_\varepsilon\|_K,$$

that is, there is no contribution to the error from cells in the interior of Ω. Hence, whatever the form of the right-hand side f, the optimal strategy is to refine the elements adjacent to the boundary and to leave the others unchanged. This requires, of course, that the function f is integrated exactly in computing the components of the load vector.

Example 2: Computation of point stresses. Next, we consider the Poisson problem (3.2) on the square domain $\Omega := (-1, 1)^2$. For a smooth solution, we want to compute $J(u) = \partial_1 u(0)$. By *a priori* analysis, we know that, for our first-degree finite element approximation on quasi-uniform meshes, we have

$$|\partial_1 e(0)| = \mathcal{O}(h) \quad (h \to 0), \tag{3.20}$$

provided that the solution has bounded second derivatives $\nabla^2 u$ (see Rannacher and Scott (1982)). Hence, in order to achieve a solution accuracy TOL, we need (in two dimensions) a mesh with $N \approx h^{-2} \approx \mathrm{TOL}^{-2}$ cells. We will examine what we can achieve on meshes constructed on the basis of the weighted *a posteriori* error estimate (3.18). Also in the present case, the adjoint solution does not exist in the sense of $H_0^1(\Omega)$, such that for practical use we have to regularize the functional, for instance taking

$$J_\varepsilon(u) = |B_\varepsilon|^{-1} \int_{B_\varepsilon} \partial_1 u \, \mathrm{d}x = \partial_1 u(0) + \mathcal{O}(\varepsilon^2),$$

where $B_\varepsilon = \{x \in \Omega, |x| < \varepsilon\}$, and $\varepsilon = \mathrm{TOL}$. The corresponding adjoint solution z behaves like $|\nabla^2 z(x)| \approx d(x)^{-3}$, where $d(x) = |x| + \varepsilon$. Using this *a priori* information and the local interpolation estimate (3.8), we can bound the weights ω_K in the *a posteriori* error estimate (3.18) as follows:

$$\omega_K \le c_i h_K^2 \|\nabla^2 z\|_K \le \tilde{c}_i h_K^3 d_K^{-3},$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Table 3.1. Computing $\partial_1 u(0)$ using the weighted estimator $\eta_\omega(u_h)$ ($L = \#$ levels)

TOL	N	L	$\lvert J_\varepsilon(e)\rvert$	$\eta_\omega(u_h)$
4^{-3}	940	9	4.10e-1	1.42e-2
4^{-4}	4912	12	4.14e-3	3.50e-3
4^{-5}	20980	15	2.27e-4	9.25e-4
4^{-6}	86740	17	5.82e-5	2.38e-4

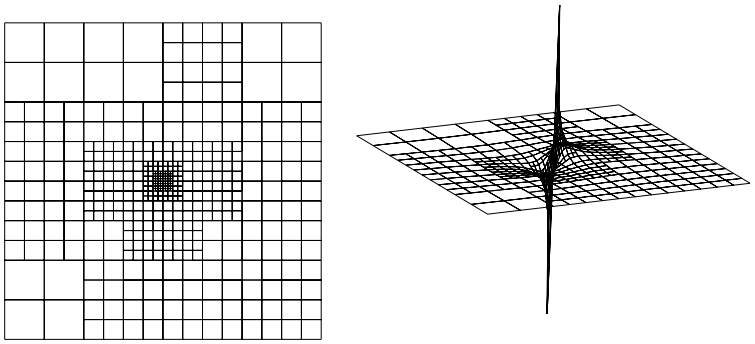


Fig. 3.1. Refined mesh and approximate adjoint solution for computing $\partial_1 u(0)$ using the *a posteriori* error estimator $\eta_\varepsilon(u_h)$, with $\varepsilon = \mathrm{TOL} = 4^{-4}$

where $d_K := \min_{x\in K} |d(x)|$. This leads us to the *a posteriori* error estimate

$$|J_\varepsilon(e)| \le \eta_\omega(u_h) := \tilde{c}_i \sum_{K\in\mathbb{T}_h} \rho_K h_K^3 d_K^{-3}, \tag{3.21}$$

which can be used for successive mesh refinement by different strategies which will be discussed in detail in Section 5, below. Here, we follow the so-called 'error-balancing strategy', by which the mesh is adapted such that, for the given tolerance TOL, the 'error indicators' $\eta_K := \rho_K h_K^3 d_K^{-3}$ are equilibrated (as well as possible) over the mesh $\mathbb{T}_h$ according to

$$\eta_K \approx \frac{\mathrm{TOL}}{N}, \quad N = \#\{K \in \mathbb{T}_h\}. \tag{3.22}$$

Results obtained in this way for a sequence of error tolerances TOL are shown in Table 3.1, while Figure 3.1 displays the balanced mesh for $\mathrm{TOL} = 4^{-4}$ and the approximation to the adjoint solution z_ε, computed on this mesh

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

(see Becker and Rannacher (1996*b*)). In order to interpret Table 3.1, we supply the following argument. Localization of the *a priori* error estimate (3.20) suggests that

$$\rho_K = \mathcal{O}(h_K) \quad (\mathrm{TOL} \to 0),$$

a (heuristic) assumption which will be addressed in some more detail in Section 5, below. Under this assumption the above error estimate becomes

$$|J_\varepsilon(e)| \;\approx\; \tilde{c}_i \sum_{K\in\mathbb{T}_h} h_K^4 d_K^{-3}. \tag{3.23}$$

Let us assume that the mesh has been adapted according to (3.22):

$$\eta_K \approx \frac{h_K^4}{d_K^3} \approx \frac{\mathrm{TOL}}{N}.$$

From this, we derive

$$h_K^2 \approx d_K^{3/2}\left(\frac{\mathrm{TOL}}{N}\right)^{1/2},$$

and consequently,

$$N = \sum_{K\in\mathbb{T}_h} h_K^2 h_K^{-2} = \left(\frac{N}{\mathrm{TOL}}\right)^{1/2} \sum_{K\in\mathbb{T}_h} h_K^2 d_K^{-3/2} \approx \left(\frac{N}{\mathrm{TOL}}\right)^{1/2}.$$

This implies that $N \approx \mathrm{TOL}^{-1}$, which is better than the $N \approx \mathrm{TOL}^{-2}$ that could be achieved on uniformly refined meshes. This predicted asymptotic behaviour is well confirmed by the results shown in Table 3.1. We emphasize that in this example strong 'mesh refinement' occurs, though the solution is smooth. In fact, this phenomenon should rather be interpreted as 'mesh coarsening' away from the point of evaluation.

Remark 3.6. In the two preceding examples the adjoint solution z could be described analytically, leading to *a priori* bounds for the weights ω_K in the *a posteriori* error estimates. In practice, this will usually not be the case and information about z has to be provided numerically by solving the adjoint problem. Strategies for the computational evaluation of the *a posteriori* error estimates will be discussed in Section 5, below.

3.2. Approximation of a symmetric eigenvalue problem

We present a simple application of the optimal control approach laid out in Proposition 2.1. Consider the first eigenvalue problem of the Laplacian,

$$-\Delta u = \lambda u \quad \text{in } \Omega, \quad u = 0 \;\text{ on } \partial\Omega, \tag{3.24}$$

on a bounded domain $\Omega \subset \mathbf{R}^d$. Computing the smallest eigenvalue and the corresponding eigenfunction of this problem may be written in variational

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

form: we seek $u \in V := H_0^1(\Omega)$ such that

$$A(u,u) = \min_{v \in V} \{A(v,v) : \|v\|^2 = 1\}, \tag{3.25}$$

where again $A(u,\varphi) = (\nabla u, \nabla \varphi)$. We call problem (3.24) 'H^2-regular' if eigenfunctions $u \in V$, $\|u\| = 1$, are also in $H^2(\Omega)$ and satisfy the *a priori* bound

$$\|\nabla^2 u\| \le c_s \|\Delta u\| = c_s \lambda. \tag{3.26}$$

The constrained optimization problem (3.25) is solved by the Lagrangian approach. Introducing the Lagrangian functional defined on pairs $x = \{u, \lambda\} \in X := V \times \mathbf{R}$,

$$L(x) := A(u,u) - \lambda\{\|u\|^2 - 1\},$$

the solution $x = \{u, \lambda\} \in X$ of (3.24) is a stationary point of $L(\cdot)$,

$$L'(x;y) = 0, \quad \text{for all } y = \{\varphi, \mu\} \in X, \tag{3.27}$$

with the derivative functional

$$L'(x;y) := 2A(u,\varphi) - 2\lambda(u,\varphi) + \mu\{\|u\|^2 - 1\}.$$

For discretizing this problem, we choose finite element subspaces $V_h \subset V$ as described above and set $X_h := V_h \times \mathbf{R}$. The approximations $x_h = \{u_h, \lambda_h\} \in X_h$ are determined by the finite-dimensional system

$$L'(x_h; y_h) = 0, \quad \text{for all } y_h \in X_h, \tag{3.28}$$

or, in explicit form,

$$A(u_h, \varphi_h) = \lambda_h (u_h, \varphi_h), \quad \text{for all } \varphi_h \in V_h, \quad \|u_h\|^2 = 1. \tag{3.29}$$

Using Proposition 2.1, we will deduce the following result.

Proposition 3.2. Suppose that, for certain solutions u and u_h of the eigenvalue problems (3.24) and (3.29), we have

$$\|u - u_h\| \le 2, \tag{3.30}$$

and that the problem is H^2-regular. Then we have the *a posteriori* error estimate

$$0 \;\le\; \lambda_h - \lambda \;\le\; 2c_i c_s \lambda_h \left(\sum_{K \in \mathbb{T}_h} h_K^4 \rho_K^2 \right)^{1/2}, \tag{3.31}$$

with the cell-wise residual terms

$$\rho_K := \|R(u_h, \lambda_h)\|_K + h_K^{-1/2} \|r(u_h)\|_{\partial K},$$

where the cell residual is $R(u_h, \lambda_h) := \Delta u_h + \lambda_h u_h$, and the edge residual $r(u_h)$ is defined as before.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Proof. **(i)** For any solution $x = \{u, \lambda\} \in X$ of (3.27) and any solution $x_h = \{u_h, \lambda_h\} \in X_h$ of (3.28),

$$\lambda = J(u) := A(u,u), \qquad \lambda_h = J(u_h) := A(u_h, u_h),$$

and, moreover,

$$0 \le J(u_h) - J(u) = L(x_h) - L(x).$$

From Proposition 2.2, we have for the error $e := x - x_h$ the identity

$$L(x_h) - L(x) = -\tfrac{1}{2} \min_{y_h \in X_h} L'(x_h; x - y_h) - R,$$

with the remainder

$$R = \tfrac{1}{2} \int_0^1 L'''(x_h + se; e, e, e)\, s(s-1)\, \mathrm{d}s.$$

The residual term on the right can be expressed in the form

$$\tfrac{1}{2} L'(x_h; x - y_h) = A(u_h, u - \varphi_h) - \lambda_h(u_h, u - \varphi_h) =: \rho(u_h, \lambda_h; u - \varphi_h).$$

Further, observing that the Lagrangian functional is quadratic in u and linear in λ, the third derivative $L'''(\cdot;\cdot,\cdot,\cdot)$ consists of only three nonzero terms:

$$L'''_{uu\lambda} = L'''_{u\lambda u} = L'''_{\lambda uu} = -2\mu(\psi, \varphi).$$

Accordingly, the remainder term becomes

$$R(x, x_h; e, e) = \tfrac{1}{4}(\lambda_h - \lambda)\|u - u_h\|^2.$$

This implies that

$$0 \le \lambda_h - \lambda = \rho(u_h, \lambda_h; u - \varphi_h) + \tfrac{1}{4}(\lambda_h - \lambda)\|u - u_h\|^2,$$

and consequently, in view of assumption (3.30),

$$0 < \lambda_h - \lambda \le 2 \min_{\varphi_h \in H_h} |\rho(u_h, \lambda_h; u - \varphi_h)|. \tag{3.32}$$

(ii) Splitting the integrals on the right in (3.32) into their contributions from the cells $K \in \mathbb{T}_h$ and integrating cell-wise by parts, we obtain similarly

$$\begin{aligned} |\rho(u_h, \lambda_h; u - \varphi_h)| &= \Big| \sum_{K \in \mathbb{T}_h} \big\{ (R(u_h, \lambda_h), u - \varphi_h)_K - (r(u_h), u - \varphi_h)_{\partial K} \big\} \Big| \\ &\le \sum_{K \in \mathbb{T}_h} \rho_K \omega_K, \end{aligned}$$

with the residual terms

$$\rho_K := \|R(u_h, \lambda_h)\|_K + h_K^{-1/2} \|r(u_h)\|_{\partial K},$$

and the weights

$$\omega_K := \|u - \varphi_h\|_K + h_K^{1/2} \|u - \varphi_h\|_{\partial K}.$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

(iii) By the interpolation estimate (3.7), we have

$$\min_{\varphi_h \in V_h} \left(\sum_{K \in \mathbb{T}_h} h_K^{-4} \omega_K^2 \right)^{1/2} \leq c_i \|\nabla^2 u\|. \tag{3.33}$$

This implies for the term in (3.32) that

$$\min_{\varphi_h \in V_h} |A(u_h, u-\varphi_h) - \lambda_h(u_h, u-\varphi_h)| \leq c_i \left(\sum_{K \in \mathbb{T}_h} h_K^4 \rho_K^2 \right)^{1/2} \|\nabla^2 u\|.$$

Hence, by the H^2-regularity (3.26) and observing $\lambda \leq \lambda_h$, it follows that

$$\lambda_h - \lambda \leq 2 c_i c_s \lambda_h \left(\sum_{K \in \mathbb{T}_h} h_K^4 \rho_K^2 \right)^{1/2},$$

which completes the proof. □

Notes and references

Estimate (3.31) was derived by Larson (2000), by direct computation. Nystedt (1995) gives the alternative error bound for the eigenvalues

$$\lambda_h - \lambda \leq c \lambda_h \sum_{K \in \mathbb{T}_h} h_K^2 \rho_K^2, \tag{3.34}$$

which reflects the dependence of the eigenvalue error on the square of the energy norm error of the eigenfunctions known from *a priori* analysis. The approach described above has been extended to *non-selfadjoint* eigenvalue problems in Heuveline and Rannacher (2001*a*). Its application for eigenvalue problems in hydrodynamic stability theory is developed in Heuveline and Rannacher (2001*b*).

3.3. A linear transport problem

The previous examples illustrate the DWR method applied to elliptic problems in which error propagation is isotropic. Next, we consider the other extreme of uni-directional error transport, as present in linear transport problems of the form

$$\beta \cdot \nabla u = f \quad \text{in } \Omega, \quad u = g \ \text{ on } \Gamma_-. \tag{3.35}$$

Here, $\Gamma_- = \{x \in \partial\Omega,\, n \cdot \beta < 0\}$ is the 'inflow boundary' and $\Gamma_+ = \partial\Omega \setminus \Gamma_-$ the 'outflow boundary'. The natural solution space is $V := \{v \in L^2(\Omega) : \beta \cdot \nabla v \in L^2(\Omega)\}$. This problem is discretized by the Galerkin finite element with streamline diffusion stabilization called SDFEM (see Hughes and Brooks (1982) and Johnson (1987)). The discretization is based on

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

first-degree polynomials on regular quadrilateral meshes $\mathbb{T}_h = \{K\}$ as described above:

$$V_h = \{v \in H^1(\Omega),\, v_{|K} \in \tilde{Q}_1(K),\, K \in \mathbb{T}_h\} \subset V.$$

The discrete solution $u_h \in V_h$ is defined by

$$(\beta{\cdot}\nabla u_h, \Phi_h) - (\beta_n u_h, \varphi_h)_- = (f, \Phi_h) - (\beta_n g, \varphi_h)_-, \quad \text{for all } \varphi_h \in V_h, \tag{3.36}$$

where $\Phi_h := \varphi_h + \delta\beta{\cdot}\nabla\varphi_h$, and $\beta_n = \beta{\cdot}n$, while the parameter function δ is determined locally by $\delta_K = \kappa h_K$ (assuming that $|\beta| = 1$). In the formulation (3.36) the inflow boundary condition is imposed in the weak sense. The right-hand and left-hand side of (3.36) define a bilinear form $A_h(\cdot,\cdot)$ and a linear form $F_h(\cdot)$, respectively. Using this notation, (3.36) may be written as

$$A_h(u_h, \varphi_h) = F_h(\varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{3.37}$$

The corresponding residual is $\rho_h(u_h,\cdot) := F_h(\cdot) - A_h(u_h,\cdot)$.

Let $J(\cdot)$ be a given functional defined on V for controlling the error $e = u - u_h$. Following our general approach, we consider the corresponding adjoint problem

$$A_h(\varphi, z) = J(\varphi), \quad \text{for all } \varphi \in V, \tag{3.38}$$

which is a (generalized) transport problem with transport in the negative β-direction. We note that here we use the stabilized bilinear form $A_h(\cdot,\cdot)$ in the duality argument, in order to achieve an optimal treatment of the stabilization terms. The existence of the adjoint solution $z \in V$ follows by standard variational arguments. Then, the general result of Proposition 2.6 yields the error representation

$$J(e) \,=\, \rho_h(u_h, z{-}\varphi_h) \,=\, (\beta{\cdot}\nabla e, z{-}\varphi_h + \delta\beta{\cdot}\nabla(z{-}\varphi_h)) - (\beta_n e, z{-}\varphi_h)_-,$$

for arbitrary $\varphi_h \in V_h$. This results in the following *a posteriori* error estimate.

Proposition 3.3. (Rannacher 1998*b*) For the approximation of the linear transport equation by the SDFEM, we have the *a posteriori* error estimate

$$|J(e)| \le \eta_\omega(u_h) := \sum_{K\in\mathbb{T}_h} \{\rho_K^1\omega_K^1 + \rho_K^2\omega_K^2\}, \tag{3.39}$$

where the residuals and weights are defined by

$$\rho_K^1 := \|f - \beta{\cdot}\nabla u_h\|_K, \quad \omega_K^1 := \|z{-}\varphi_h\|_K + \delta_K\|\beta{\cdot}\nabla(z{-}\varphi_h)\|_K,$$
$$\rho_K^2 := h_K^{-1/2}\|\beta_n(u_h{-}g)\|_{\partial K\cap\Gamma_-}, \quad \omega_K^2 := h_K^{1/2}\|z{-}\varphi_h\|_{\partial K\cap\Gamma_-},$$

with a suitable approximation $\varphi_h \in V_h$ to z.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

This *a posteriori* error bound explicitly contains the mesh size h_K and the stabilization parameter δ_K as well. This gives us the possibility of simultaneously adapting both parameters, which may be particularly advantageous for capturing sharp layers in the solution.

Examples. We want to illustrate the error estimator (3.39) by two thought experiments. Let $\beta \equiv$ const and $f = 0$. First, we take the functional $J(e) := (1, \beta_n e)_+$. The corresponding adjoint solution is $z \equiv 1$, so that $J(e) = 0$. This reflects the global conservation property of the SDFEM. Next, we set

$$J(e) := (1, e) + (1, \delta\beta_n e)_+.$$

The corresponding adjoint problem reads

$$(-\beta \cdot \nabla z, \varphi - \delta\beta \cdot \nabla\varphi) + (\beta_n z, \varphi)_+ = (1, \varphi) + (1, \delta\beta_n\varphi)_+.$$

Assuming that $\delta \equiv$ const, this adjoint problem has the same solution as

$$-\beta \cdot \nabla z = 1 \quad \text{in } \Omega, \quad z = \delta \ \text{ on } \Gamma_-.$$

Consequently, z is linear almost everywhere, that is, the weights in the *a posteriori* bound (3.39) are nonzero only along the lines of discontinuity. Therefore, the mesh refinement will be restricted to these critical regions although the cell residuals $\rho_K(u_h)$ may be nonzero everywhere.

Numerical test (*from Hartmann* (200x))
We consider the model problem (3.35) on the unit square $\Omega = (0,1)\times(0,1) \subset \mathbb{R}^2$ with the right-hand side $f \equiv 0$, the (constant) transport coefficient $\beta = (-1, -0.5)^T$, and the inflow data $g = 1$ on Γ_1 and $g = 0$ elsewhere. The quantity to be computed is part of the outflow, as shown in Figure 3.2:

$$J(u) := \int_{\Gamma_2} \beta\cdot nu \, \mathrm{d}s.$$

The corresponding meshes and the solutions u_h and z_h are shown in Figure 3.2. There is no mesh refinement along the upper line of discontinuity of the solution since here the cell residuals of the adjoint solution z_h are almost zero (z is constant).

Notes and references
Here we have considered only the adaptive SDFEM for simple *linear and scalar* transport problems (for more examples see Houston, Rannacher and Süli (2000)). Analogous results have also been obtained by the discontinuous Galerkin finite element 'dG(r) method' (for a comparison see Hartmann (2000)). Earlier work on the adaptive SDFEM for linear as well as nonlinear problems (*e.g.*, Burgers' equation and Euler equations) using duality

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

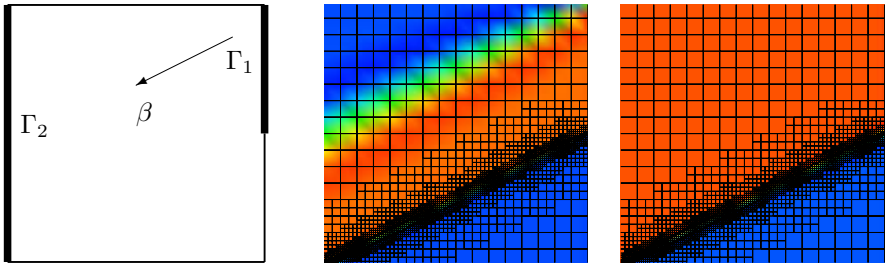


Fig. 3.2. Configuration and results for the model transport problem (3.35): primal solution (centre) and dual solution (right) on an adaptively refined mesh

and 'stability constants' is Johnson (1990), Hansbo and Johnson (1991), Eriksson and Johnson (1993), Eriksson, Johnson and Larsson (1998), and Johnson and Szepessy (1995). The DWR approach in combination with the SDFEM and the dG(r) method for nonlinear problems has been considered by Führer (1996), Führer and Rannacher (1997), and Hartmann (200x). An application to transport equations with stiff source terms is described in Hebeker and Rannacher (1999).

4. Norm-based error estimation

In this section, we discuss techniques for deriving *a posteriori* estimates for the error in global norms and also with respect to local output quantities which use the concept of 'energy error estimation'. For simplicity, we continue using the Poisson problem as our prototype.

4.1. A posteriori error bounds in global norms

By the same type of argument as used in Proposition 3.1, we can also derive the traditional global error estimates in the energy and the L^2 norm.

(i) Energy error bound. First, we use the functional

$$J(\varphi) = \|\nabla e\|^{-1}(\nabla e, \nabla \varphi)$$

in the adjoint problem. Its solution $z \in V$ satisfies $\|\nabla z\| \leq 1$. We obtain the estimate

$$\|\nabla e\| \leq \sum_{K\in\mathbb{T}_h} \rho_K \, \omega_K \leq \left(\sum_{K\in\mathbb{T}_h} h_K^2 \rho_K^2 \right)^{1/2} \left(\sum_{K\in\mathbb{T}_h} h_K^{-2} \omega_K^2 \right)^{1/2},$$

with residual terms and weights as defined above. Now, we use an extension

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

of the interpolation estimate (3.7),

$$\left(\sum_{K\in\mathbb{T}_h} \left\{ h_K^{-2} \|z - \tilde{\imath}_h z\|_K^2 + h_K^{-1} \|z - \tilde{\imath}_h z\|_{\partial K}^2 \right\} \right)^{1/2} \leq \tilde{c}_i \, \|\nabla z\|, \tag{4.1}$$

where $\tilde{\imath}_h z \in V_h$ is a modified nodal interpolation which is defined and stable on $H^1(\Omega)$ (for such a construction see, *e.g.*, Brenner and Scott (1994)). This gives us

$$\|\nabla e\| \leq \tilde{c}_i \left(\sum_{K\in\mathbb{T}_h} h_K^2 \rho_K^2 \right)^{1/2} \|\nabla z\|.$$

Finally, observing the *a priori* bound for $\|\nabla z\|$, we conclude the *a posteriori energy error* estimate

$$\|\nabla e\| \leq \eta_E(u_h) = \tilde{c}_i \left(\sum_{K\in\mathbb{T}_h} h_K^2 \rho_K^2 \right)^{1/2}. \tag{4.2}$$

(ii) L^2-norm error bound. Next, we use the functional

$$J(\varphi) = \|e\|^{-1}(e, \varphi)$$

in the adjoint problem. If the (polygonal) domain Ω is convex, the dual solution $z \in V$ is in $H^2(\Omega)$ and admits the *a priori* bound $c_s := \|\nabla^2 z\| \leq 1$. As in (i), this yields the error estimate

$$\|e\| \leq \left(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2 \right)^{1/2} \left(\sum_{K\in\mathbb{T}_h} h_K^{-4} \omega_K^2 \right)^{1/2}.$$

Now we use the stronger version of the interpolation estimate (4.1),

$$\left(\sum_{K\in\mathbb{T}_h} \left\{ h_K^{-4} \|z - i_h z\|_K^2 + h_K^{-3} \|z - i_h z\|_{\partial K}^2 \right\} \right)^{1/2} \leq c_i \, \|\nabla^2 z\|, \tag{4.3}$$

to obtain

$$\|e\| \leq c_i \left(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2 \right)^{1/2} \|\nabla^2 z\|.$$

Finally, observing the *a priori* bound for $\|\nabla^2 z\|$, we conclude the L^2-norm *a posteriori* error estimate

$$\|e\| \leq \eta_{L^2}(u_h) = c_i c_s \left(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2 \right)^{1/2}. \tag{4.4}$$

In this estimate, the information about the global error sensitivities contained in the dual solution is condensed into just one 'stability constant'

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

$c_s = \|\nabla^2 z\|$. This is appropriate in estimating the global L^2 error in the case with constant diffusion coefficient. For more general situations with strongly varying coefficients, it may be advisable rather to follow the DWR approach in which the information from the dual solution is kept within the *a posteriori* error estimator as weights:

$$\|e\| \le c_i \sum_{K\in\mathbb{T}_h} h_K^2 \rho_K \|\nabla^2 z\|_K. \tag{4.5}$$

Below, we will present an example to support this claim.

4.2. Application for L^2-error estimation

We consider the finite element approximation of a diffusion equation with variable coefficient,

$$-\nabla\cdot\{a\nabla u\} = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega. \tag{4.6}$$

The error is to be estimated in the L^2 norm. By this, we want to demonstrate that the use of 'weighted' *a posteriori* error estimates as proposed by the DWR method (using the approximations described above) can also be beneficial for estimating the error in a global norm. The adjoint problem for estimating the L^2 error is

$$-\nabla\cdot\{a\nabla z\} = \|e\|^{-1} e \quad \text{in } \Omega, \quad z = 0 \quad \text{on } \partial\Omega. \tag{4.7}$$

The *a posteriori* error estimates (4.4) and (4.5) have natural generalizations to the present situation, that is, the usual L^2-error estimate

$$\|e\| \le \eta_{L^2}(u_h) := c_i c_s \left(\sum_{K\in\mathbb{T}_h} h_K^4 \rho_K^2 \right)^{1/2}, \quad c_s := \|\nabla^2 z\|, \tag{4.8}$$

and the 'weighted' L^2-error estimate

$$\|e\| \le \eta^{\omega}_{L^2}(u_h) := c_i \sum_{K\in\mathbb{T}_h} h_K^2 \rho_K \omega_K, \quad \omega_K := \|\nabla^2 z\|_K, \tag{4.9}$$

with the cell residuals $\rho_K := \|R(u_h)\|_K + h_K^{-1/2}\|r(u_h)\|_{\partial K}$ as defined in Remark 3.5. Both error estimators are evaluated by replacing the second derivatives of the dual solution z by second-order difference quotients of an approximation $z_h \in V_h$, as described above:

$$\omega_K \approx \tilde\omega_K := \|\nabla_h^2 z_h\|_K, \quad c_s \approx \tilde c_s := \left(\sum_{K\in\mathbb{T}_h} \|\nabla_h^2 z_h\|_K^2 \right)^{1/2}. \tag{4.10}$$

The interpolation constant is set to $c_i = 0.2$. The functional $J(\cdot)$ is evaluated by replacing the unknown solution u by a patch-wise higher-order interpolation $I_h^{(2)} u_h$ of the computed approximation u_h, that is, $e \approx I_h^{(2)} u_h - u_h$.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

The quality of the resulting two *a posteriori* error estimators $\tilde{\eta}_{L^2}(u_h)$ and $\tilde{\eta}^{\omega}_{L^2}(u_h)$ for mesh adaptation will be compared below for a representative model case.

Numerical test

We choose the square domain $\Omega = (-1,1) \times (-1,1)$ and the nonconstant coefficient function $a(x) = 0.1 + e^{3(x_1+x_2)}$ with right-hand side $f \equiv 0.1$. A reference solution is generated by a computation on a very fine mesh. The meshes are refined according to the 'error-balancing strategy' by balancing (as well as possible) the cell-wise indicators $\eta_K := h_K^4 \rho_K^2$ or $\eta_K := h_K^2 \rho_K \tilde{\omega}_K$ over the mesh $\mathbb{T}_h$. Figure 4.1 shows error plots and meshes obtained by the two different strategies.

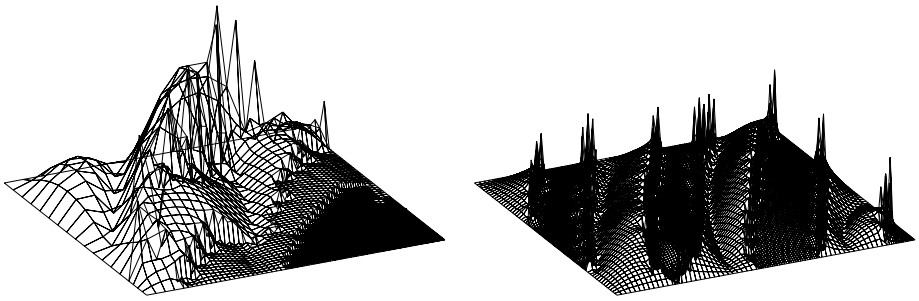


Fig. 4.1. Errors obtained by η_{L^2} (left, scaled by 1/30) and $\eta^{\omega}_{L^2}$ (right, scaled by 1/10) on meshes with $N \sim 10000$ cells

We see that carrying the weights $\|\nabla^2 z\|_K$ within the estimator yields a distribution of mesh cells that is much better adapted to the solution structure determined by the nonconstant coefficient $a(x)$. Condensing all the weighting terms into just one global 'stability constant' $c_s \approx \|\nabla^2 z\|_\Omega$ loses this detail information. Further, the results in Table 4.1 show that the error bounds resulting from the 'weighted' *a posteriori* error estimates are more accurate.

4.3. A posteriori error estimates for functional output based on energy error estimates

Several other approaches to 'goal-oriented' *a posteriori* error control have been proposed in the literature which, like the DWR method, use duality arguments for error representation but employ global 'energy norm' error estimates. We continue to use the notation of the preceding section particularly that of the model problem (3.2). Let $J(\cdot)$ again be a (linear) error

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Table 4.1. Results obtained by the global L^2-error estimator $\tilde{\eta}_{L_2}$ (top) compared to the weighted estimator $\tilde{\eta}^{\omega}_{L_2}$ (bottom) ($I_{\text{eff}} := \eta(u_h)/\|e\|$, $L = \#$ refinement levels)

TOL	N	L	$\|e\|$	$\tilde{\eta}_{L^2}$	I_{eff}	$\tilde{c}_s$
4^{-2}	2836	9	6.40e-2	2.32e-1	3.62	3.02
4^{-3}	5884	10	2.13e-2	1.21e-1	5.68	3.26
4^{-4}	15736	11	7.36e-3	4.76e-2	6.46	3.55
4^{-5}	23380	11	5.59e-3	3.12e-2	5.58	3.39

TOL	N	L	$\|e\|$	$\tilde{\eta}^{\omega}_{L^2}$	I_{eff}	$\tilde{c}_s$
4^{-4}	220	5	6.77e-2	5.24e-2	0.77	.....
4^{-5}	592	6	2.21e-2	2.59e-2	1.17	.....
4^{-6}	892	6	1.19e-2	1.54e-2	1.29	.....
4^{-7}	2368	7	5.11e-3	7.17e-3	1.40	.....

functional, $z \in V$ the associated dual solution and $z_h \in V_h$ its finite element projection. For the errors $e := u - u_h$ and $e^* := z - z_h$, by Galerkin orthogonality, we have

$$J(e) = A(e, z) = A(e, e^*) = A(u, e^*) = (f, e^*). \tag{4.11}$$

We recall the definition of the corresponding residual functionals

$$\rho(u_h, \varphi) := (f, \varphi) - A(u_h, \varphi), \qquad \rho^*(z_h, \varphi) := J(\varphi) - A(\varphi, z_h).$$

Clearly, these residuals vanish for $\varphi_h \in V_h$. Next, a suitable enlarged space $\hat{V}_h \supset V_h$ is introduced, for instance, by adding certain higher-order polynomials in each cell (see again Ainsworth and Oden (1993, 1997*a*)). The functions in $\hat{V}_h$ are allowed to be discontinuous. We assume that the residual functionals ρ and ρ^* possess proper extensions $\hat{R}$ and $\hat{R}^*$ to $\hat{V}_h$ (involving, for example, local stress averaging). The natural 'cell-wise' defined extension of the bilinear form $A(\cdot, \cdot)$ is denoted by $A_h(\cdot, \cdot)$. With this notation approximations $\hat{e}$ and $\hat{e}^* \in \hat{V}_h$ are generated for the errors e and e^* by solving the defect equations

$$\begin{aligned} A_h(\hat{e}, \hat{\varphi}_h) &= \hat{\rho}(u_h, \hat{\varphi}_h), \quad \text{for all } \hat{\varphi}_h \in \hat{V}_h, \\ A_h(\hat{\varphi}_h, \hat{e}) &= \hat{\rho}^*(z_h, \hat{\varphi}_h), \quad \text{for all } \hat{\varphi}_h \in \hat{V}_h. \end{aligned} \tag{4.12}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Owing to the allowed discontinuity of functions in $\hat{V}_h$, solving (4.12) usually reduces to solving local defect problems separately on each mesh cell $K \in \mathbb{T}_h$. Now, the assumption is made that the space $\hat{V}_h$ is sufficiently large that

$$e \in \hat{V}_h. \tag{4.13}$$

Of course, in general, this condition cannot be satisfied exactly by a finite-dimensional space $\hat{V}_h$. Therefore, condition (4.13) renders all the resulting conclusions heuristic. Nevertheless, assuming (4.13) to be true, we can take e as a test function in (4.12), obtaining

$$A_h(\hat{e}, e) = \hat{\rho}(u_h, e) = \rho(u_h, e) = \|\nabla e\|^2,$$

$$A_h(\hat{e}^*, e) = \hat{\rho}^*(z_h, e) = \rho^*(z_h, e) = J(e).$$

This trivially implies that $\|\nabla e\| \leq \|\nabla \hat{e}\|_h$, and consequently,

$$|J(e)| \;=\; |A_h(e, \hat{e}^*)| \;\leq\; \|\nabla e\|\,\|\nabla \hat{e}^*\| \;\leq\; \|\nabla \hat{e}\|\,\|\nabla \hat{e}^*\|. \tag{4.14}$$

By an elementary argument, this error bound can be improved to

$$|J(u) - \hat{J}(u_h)| \;\leq\; \tfrac{1}{2}\|\nabla \hat{e}\|\,\|\nabla \hat{e}^*\|, \tag{4.15}$$

where $\hat{J}(u_h) := J(u_h) + \frac{1}{2}A_h(\hat{e}, \hat{e}^*)$ (Machiels *et al.* 1998). The error bound (4.14) requires precise energy-norm estimates for e as well as e^*, which are difficult to achieve for singular z, such as, for example, in point error estimation. Further, (4.14) does not directly yield criteria for local mesh refinement. In order to obtain those, one may return to the very basic error representation (4.11) which directly implies the error bound

$$|J(e)| \;\leq\; \sum_{K \in \mathbb{T}_h} \|\nabla e\|_K \|\nabla e^*\|_K.$$

This is then replaced heuristically by

$$|J(e)| \;\approx\; \sum_{K \in \mathbb{T}_h} \|\nabla \hat{e}\|_K \|\nabla \hat{e}^*\|_K. \tag{4.16}$$

This last step lacks justification (and may even lead to a systematically wrong result) since the local total error $\|\nabla e^*\|_K$ is not necessarily bounded by the locally obtained approximation $\|\nabla \hat{e}^*\|_K$ due to possible error pollution. This defect may not be critical in generating good meshes for elliptic model problems but can be disastrous in situations where strong error transport occurs. The DWR method avoids this problem by correctly replacing the terms $\|\nabla e\|_K \|\nabla e^*\|_K$ by $\rho_K(u_h)\omega_K(z)$ where the local residual $\rho_K(u_h)$ can be evaluated exactly and the local weights $\omega_K(z) \sim \|z - i_h z\|_K$ *do not* suffer from error pollution provided, of course, that z is computed with sufficient accuracy. Alternatively, it has been suggested that we convert the globally multiplicative *a posteriori* estimate (4.14) into a locally additive

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

one similar to (4.20):

$$|J(e)| \leq \sum_{K\in\mathbb{T}_h} \{\theta\|\nabla\hat{e}\|_K^2 + \theta^{-1}\|\nabla\hat{e}^*\|_K^2\}, \tag{4.17}$$

for some $\theta > 0$. Again, because of the necessarily global choice of θ, this procedure does not allow us to balance the cell-wise contributions of $\hat{e}$ and $\hat{e}^*$ properly, which is critical in the case of strongly differing behaviour of u and z.

For completeness, we briefly sketch a variant of the above approach, which requires upper *and* lower bounds for the energy-norm error and fully exploits the fact that the energy bilinear form is a scalar product. Suppose that we have available approximations $\hat{e}$ and $\check{e}$ for the error e which yield upper as well as lower bounds for the energy error:

$$\|\nabla\underline{e}\| \leq \|\nabla e\| \leq \|\nabla\hat{e}\|. \tag{4.18}$$

Starting from the error representation (4.11), the idea is to use the parallelogram identity for scalar products to obtain

$$J(e) = A(\theta e, \theta^{-1}e^*) = \tfrac{1}{4}\|\nabla(\theta e+\theta^{-1}e^*)\|^2 - \tfrac{1}{4}\|\nabla(\theta e-\theta^{-1}e^*)\|^2,$$

with a suitable balancing parameter $\theta > 0$. In view of (4.18), this implies the following *a posteriori* error estimate:

$$\check{\eta}_E(u_h, z_h) \leq J(e) \leq \hat{\eta}_E(u_h, z_h), \tag{4.19}$$

with the lower and upper bounds given by

$$\check{\eta}_E(u_h, z_h) := \tfrac{1}{4}\|\nabla(\theta\check{e}+\theta^{-1}\check{e}^*)\|^2 - \tfrac{1}{4}\|\nabla(\theta\hat{e}-\theta^{-1}\hat{e}^*)\|^2,$$

$$\hat{\eta}_E(u_h, z_h) := \tfrac{1}{4}\|\nabla(\theta\hat{e}+\theta^{-1}\hat{e}^*)\|^2 - \tfrac{1}{4}\|\nabla(\theta\check{e}-\theta^{-1}\check{e}^*)\|^2.$$

Again it is not clear how to organize effective local mesh adaptation on the basis of a global estimate of the form (4.19). The simple idea of 'localizing' the error bound $\hat{\eta}_E(u_h, z_h)$ according to

$$\hat{\eta}_E(u_h, z_h) \leq \tfrac{1}{4}\sum_{K\in\mathbb{T}_h} \left|\|\nabla(\theta\hat{e}+\theta^{-1}\hat{e}^*)\|_K^2 - \tfrac{1}{4}\|\nabla(\theta\check{e}-\theta^{-1}\check{e}^*)\|_K^2\right| \tag{4.20}$$

again suffers from the fact that the parameter θ is to be chosen globally. Hence, the multiplicative local interaction of e and e^*, which is the essential feature of the DWR method, would be lost here.

Finally, we note that the argument leading to the *a posteriori* error bound (4.14) can be generalized to nonsymmetric problems. If the symmetric energy form $A(\cdot,\cdot)$ is replaced by a nonsymmetric bilinear form, for example,

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

$A(\cdot,\cdot) := (\nabla\cdot,\nabla\cdot) + (b\cdot\nabla\cdot,\cdot)$, the error estimate (4.14) contains an additional constant measuring the deviation of $A(\cdot,\cdot)$ from symmetry:

$$|J(e)| \leq \frac{1+\lambda\mu}{2} \|\nabla\hat{e}\| \, \|\nabla\hat{e}^*\|, \tag{4.21}$$

where

$$\lambda := \inf_{v\in V} \left\{ \sup_{\varphi\in V} \frac{A(v,\varphi)}{\|\nabla v\|\|\nabla\varphi\|} \right\}, \qquad \mu := \inf_{v\in V} \left\{ \sup_{\varphi\in V} \frac{A(\varphi,v)}{\|\nabla\varphi\|\|\nabla v\|} \right\}.$$

In practice, it would be very difficult to obtain realistic bounds for the *global* stability parameters λ and μ. The DWR method is designed to avoid this critical difficulty simply by numerically solving the associated adjoint problem with higher accuracy and thereby implicitly detecting the *local* sensitivities of the problem.

Notes and references

The approaches described above were originally proposed in a series of papers by Babuška and Miller (1984*a*, 1984*b*, 1984*c*) as a post-processing tool for achieving improved accuracy. Recently, they have been extensively elaborated for several model applications in Paraschivoiu and Patera (1998), Peraire and Patera (1998), and Oden and Prudhomme (1999).

5. Evaluation of estimates and mesh adaptation

In this section, we discuss some practical aspects of the DWR method. The first question is how to evaluate the *a posteriori* error representations and estimates derived in Section 3 in practice. This particularly involves the approximation of the 'adjoint solution'. The second question concerns the use of these estimates in the course of successive mesh adaptation. Finally, in the solution of nonlinear problems we have to combine error estimation and mesh adaptation with nonlinear iteration and linearization. The latter aspect will be discussed in Section 6.

5.1. Evaluation of a posteriori error bounds

First, we will discuss the evaluation of the error representation (3.17) or the resulting error bound (3.18), again using the Poisson problem as our prototype (see Becker and Rannacher (1996*b*)). There are essentially two different approaches. In the first one the dual solution z is replaced by some approximation $\hat{z}_h$ obtained from solving the adjoint problem by a higher-order method or on a finer mesh. In the second one the quantity $z - \varphi_h$ is estimated using the finite element projection $z_h \in V_h$ of z on the same

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Table 5.1. Efficiency of the *weighted* error indicators for computing the point-error $J(e) = |e(0)|$ on uniformly refined meshes

L	N	$I_{\text{eff}}(\eta_\omega^1)$	$I_{\text{eff}}(\eta_\omega^2)$	$I_{\text{eff}}(\eta_\omega^3)$
1	4^3	6.667	12.82	2.066
2	4^4	1.253	13.51	45.45
3	4^5	1.052	3.105	9.345
4	4^6	1.007	2.053	7.042
5	4^7	1.003	1.886	6.667

mesh. As our quality measure, we use the over-estimation factor

$$I_{\text{eff}} := \Big|\frac{\eta(u_h)}{J(e)}\Big|,$$

referred to as the (reciprocal) 'effectivity index' (see Babuška and Rheinboldt (1978*b*)). Results of tests with these methods are reported in Table 5.1.

(i) Global higher-order approximation. Let $\hat{z}_h \in \hat{V}_h$ be an approximation to z obtained by a higher-order method or on a finer mesh $\hat{\mathbb{T}}_h$ with corresponding finite element space $\hat{V}_h$:

$$a(\hat{\varphi}_h, \hat{z}_h) = J(\hat{\varphi}_h), \quad \text{for all } \hat{\varphi}_h \in \hat{V}_h.$$

The additional error caused by this approximation can be described by writing the error representation (3.17) in a symmetric form. To this end, we use Galerkin orthogonality of $e := u - u_h$ and $\hat{e}^* := z - \hat{z}_h$ to obtain

$$J(e) \;=\; a(e, \hat{z}_h - \varphi_h) + a(u - \hat{\varphi}_h, \hat{e}^*),$$

with arbitrary $\varphi_h \in V_h$ and $\hat{\varphi}_h \in \hat{V}_h$. This can be rewritten as

$$\begin{aligned} J(e) \;=\; &\sum_{K\in\mathbb{T}_h} \big\{(R(u_h), \hat{z}_h - \varphi_h)_K - (r(u_h), \hat{z}_h - \varphi_h)_{\partial K}\big\} \\ &+ \sum_{K\in\hat{\mathbb{T}}_h} \big\{(u - \hat{\varphi}_h, R^*(\hat{z}_h))_K - (u - \hat{\varphi}_h, r^*(\hat{z}_h))_{\partial K}\big\}, \end{aligned} \tag{5.1}$$

with the cell and edge residuals $R(u_h)$ and $r(u_h)$ of u_h, and $R^*(\hat{z}_h)$ and $r^*(\hat{z}_h)$ of $\hat{z}_h$, respectively, defined analogously as before. The first sum on the right can easily be evaluated, while the evaluation of the second one would require generation of an improved approximation to u. Therefore, the simplest approach is to assume that this second sum is small compared

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

to the first and neglect it. The resulting error estimator reads

$$\eta_\omega^1(u_h) := \Big| \sum_{K\in\mathbb{T}_h} \big\{(R(u_h), \hat{z}_h - i_h\hat{z}_h)_K - (r(u_h), \hat{z}_h - i_h\hat{z}_h)_{\partial K}\big\}\Big|,$$

with the nodal interpolation $i_h\hat{z}_h \in V_h$. Note that we cannot simplify this approach by using $\hat{z}_h = z_h$ since then the whole error estimator would vanish. We have tested this approach using bi-quadratic approximation for $\hat{z}_h$ on the primal mesh $\mathbb{T}_h$. In the model case of the Poisson equation (3.2), it is seen by the numerical results as well as by theoretical analysis that, in this case, $\hat{\eta}_\omega(u_h)$ is asymptotically sharp, that is, $\lim_{\mathrm{TOL}\to 0} I_{\mathrm{eff}} = 1$ (see Table 5.1). However, approximating z by a higher-order finite element scheme does not seem very economical in estimating the error in the low-order scheme. A more practical alternative would be to compute an approximation to z on a sufficiently fine 'super-mesh'. Then, increasing the accuracy, arbitrarily close approximations to $J(e)$ could be obtained.

(ii) Local higher-order approximation. In most practical cases it suffices to use the finite element projection $z_h \in V_h$ of z computed on the current mesh. It is assumed that an improved approximation can be obtained by patch-wise higher-order interpolation $i_h^+ z_h$. This construction requires special care on elements with hanging nodes, in order to preserve the higher-order accuracy of the interpolation process. The resulting global error estimator reads

$$\eta_\omega^2(u_h) := \Big| \sum_{K\in\mathbb{T}_h} \big\{(R(u_h), i_h^+ z_h - z_h)_K - (r(u_h), i_h^+ z_h - z_h)_{\partial K}\big\}\Big|.$$

For bi-quadratic interpolation, we observe that $\limsup_{\mathrm{TOL}\to 0} I_{\mathrm{eff}} \approx 1.2-3$.

(iii) Approximation by difference quotients. We use the cell-wise interpolation estimate (3.7) to estimate the weights ω_K by

$$\omega_K = \|z - i_h z\|_K + h_K^{1/2}\|z - i_h z\|_{\partial K} \le c_i h_K^2 \|\nabla^2 z\|_K.$$

The second derivative $\nabla^2 z$ is then replaced by a suitable second-order difference quotient $\nabla_h^2 z_h$. A useful substitute is the approximation

$$\|\nabla^2 z\|_K \approx h_K^{-1/2}\|r(z_h)\|_{\partial K},$$

which corresponds to the edge residual $\rho_{\partial K}$ of u_h, the evaluation of which has to be implemented anyway. This yields the heuristic error estimator

$$\eta_\omega^3(u_h) := c_i \sum_{K\in\mathbb{T}_h} h_K^{-1/2}\rho_K\|r(z_h)\|_{\partial K}.$$

In this case, we observe that $\limsup_{\mathrm{TOL}\to 0} I_{\mathrm{eff}} \approx 2-10$ depending on the problem. The estimator $\eta_\omega^3(u_h)$ involves an interpolation constant c_i for

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

which sharp bounds are difficult to get. Therefore it is not suited to error estimation, but by localization it yields rather effective mesh refinement indicators $\eta_K := \rho_K \|r(z_h)\|_{\partial K}$.

The effectivity indices observed for these strategies in computing the point-value $u(0)$ (on uniformly refined meshes) are listed in Table 5.1 (from Becker and Rannacher (1996*b*)). These results indicate that, even for the simplest model situations, an asymptotic effectivity index $I_{\text{eff}} = 1$ is achievable only at the expense of high extra cost, for instance by approximating the dual solution using a higher-order method.

5.2. Strategies for mesh adaptation

Now, we discuss some popular strategies for mesh adaptation based on an *a posteriori* error estimate of the form

$$|J(e)| \leq \eta(u_h) := \sum_{K \in \mathbb{T}_h} \eta_K, \tag{5.2}$$

with certain cell-error indicators $\eta_K = \eta_K(u_h)$ obtained on the current mesh $\mathbb{T}_h$. Suppose that a tolerance TOL for the error $J(e)$ or a maximum number $N_{\max}$ of mesh cells has been prescribed. Starting from an approximate solution $u_h \in V_h$ obtained on the current mesh $\mathbb{T}_h$, the mesh adaptation may be organized by one of the following strategies.

Error-balancing strategy. Cycle through the mesh and equilibrate the local error indicators according to $\eta_K \approx \text{TOL}/N$ with $N := \#\{K \in \mathbb{T}_h\}$. This leads eventually to $\eta(u_h) \approx \text{TOL}$. Since N changes when the mesh is locally refined or coarsened, this strategy requires iteration with respect to N on each mesh level.

Fixed mesh (or error) fraction strategy. Order cells according to the size of η_K and refine a certain percentage (say 20%) of cells with largest η_K (or those which make up 70% of the estimator value $\eta(u_h)$) and coarsen those cells with smallest η_K. By this strategy, one may achieve a prescribed rate of increase of N or keep it constant within a nonstationary computation, if that is desirable.

Mesh-optimization strategy. Use a (proposed) representation

$$\eta(u_h) := \sum_{K \in \mathbb{T}_h} \eta_K \approx \int_\Omega h(x)^2 \Phi(x)\, \mathrm{d}x \tag{5.3}$$

for directly generating a formula for an optimal mesh-size distribution:

$$h_{\text{opt}}(x) = \left(\frac{W}{N_{\max}}\right)^{1/2} \Phi(x)^{-1/4}, \quad W := \int_\Omega \Phi(x)^{1/2}\, \mathrm{d}x. \tag{5.4}$$

Here, we think of a smoothly distributed mesh-size function $h(x)$ such

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

that $h_{|K} \approx h_K$. The existence of such a representation with an h-independent error density function $\Phi(x) \approx |\nabla^2 u(x)|\,|\nabla^2 z(x)|$ can be rigorously justified only under very restrictive conditions, but is generally supported by computational experience (Becker and Rannacher 1996*b*). Even for rather 'irregular' functionals $J(\cdot)$ the quantity W in (5.4) is bounded. For example, the evaluation of $J(u) = \partial_i u(P)$ for smooth u in two dimensions leads to $\Phi(x) \approx |x-P|^{-3}$ and, consequently, $W < \infty$.

Problem 5.1. Establish an asymptotic error representation of the form (5.3) on adaptively refined meshes. The essential step for this is the bound

$$\limsup_{N\to\infty} \|\nabla_h^2 u_h\|_\infty \;\le\; c\,\|\nabla^2 u\|_\infty, \tag{5.5}$$

for second-order difference quotients $\nabla_h^2 u_h$ of first-degree finite elements.

Problem 5.2. The explicit formula for $h_{\rm opt}(x)$ has to be used with care in designing a mesh. Its derivation implicitly assumes that it actually corresponds to a *scalar* mesh-size function of an isotropic mesh such that $h_{{\rm opt}|K} \approx h_K$. However, this condition is not incorporated into the formulation of the mesh-optimization problem. The treatment of anisotropic meshes containing stretched cells requires a more involved concept of mesh description and optimization.

6. Algorithmic aspects for nonlinear problems

In this section, we discuss the practical realization of the DWR method for nonlinear problems. This concerns the iterative treatment of the nonlinearity and particularly the evaluation of the *a posteriori* error estimators by linearization.

6.1. The nested solution approach

In the following, we briefly describe how the concepts of adaptivity laid out so far can be combined with an iterative solution process, for instance a Newton-type method, for general nonlinear problems as in Section 2:

$$A(u;\varphi) = F(\varphi), \quad \text{for all } \varphi \in V. \tag{6.1}$$

This inherits the basic structure of the solution processes employed in all the nonlinear applications which will be presented below. Let a desired error tolerance TOL or a maximum mesh complexity $N_{\max}$ be given. Starting from a coarse initial mesh $\mathbb{T}_0$, a hierarchy of successively refined meshes $\mathbb{T}_i$, $i \ge 1$, and corresponding finite element spaces V_i, is generated by the following algorithm.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Adaptive solution algorithm.

(0) Initialization $i = 0$: Compute an initial approximation $u_0 \in V_0$.

(1) Defect correction iteration: For $i \geq 1$, start with $u_i^{(0)} := u_{i-1} \in V_i$.

(2) Iteration step: For $j \geq 0$ evaluate the defect

$$(d_i^{(j)}, \varphi) := F(\varphi) - A(u_i^{(j)}; \varphi), \quad \varphi \in V_i. \tag{6.2}$$

Pick a suitable approximation $\tilde{A}'(u_i^{(j)}; \cdot, \cdot)$ to the derivative $A'(u_i^{(j)}; \cdot, \cdot)$ (with good stability and solubility properties) and compute a correction $v_i^{(j)} \in V_i$ from the linear equation

$$\tilde{A}'(u_i^{(j)}; v_i^{(j)}, \varphi) = (d_i^{(j)}, \varphi), \quad \text{for all } \varphi \in V_i. \tag{6.3}$$

For that, Krylov space and multigrid methods are employed using the hierarchy of already constructed meshes $\{\mathbb{T}_i, \ldots, \mathbb{T}_0\}$ (see Becker and Braack (2000)). Then, update $u_i^{(j+1)} = u_i^{(j)} + \lambda_i v_i^{(j)}$, with some relaxation parameter $\lambda_i \in (0, 1]$, set $j := j + 1$ and go back to (2). This process is repeated until a limit $\tilde{u}_i \in V_i$, is reached with a prescribed accuracy.

(3) Error estimation: Accept $\tilde{u}_i = u_i$ as the solution on mesh $\mathbb{T}_i$, solve the discrete linearized adjoint problem

$$z_i \in V_i: \quad A'(u_i; \varphi, z_i) = J'(u_i; \varphi), \quad \text{for all } \varphi \in V_i,$$

and evaluate the *a posteriori* error estimate

$$|J(u) - J(u_i)| \approx \eta_\omega(u_i). \tag{6.4}$$

For controlling the reliability of this bound, that is, the accuracy of the linearization and the determination of the approximate adjoint solutions z_i, one may use the algorithm described below. If the error estimator $\eta_\omega(u_i)$ is detected to be reliable and $\eta(u_i) \leq \text{TOL}$ or $N_i \geq N_{\max}$, then stop. Otherwise mesh adaptation yields the new mesh $\mathbb{T}_{i+1}$. Then, set $i := i+1$ and go to (1).

Remark 6.1. The defect-correction iteration (1)–(3) has to be terminated when one is close enough to the solution $u_i \in V_i$. This can be controlled by monitoring the algebraic residual $\|d_i^{(j)}\|$, which can additionally be weighted cell-wise by the current approximation z_i to the adjoint solution. However, the development of a fully satisfactory stopping criterion on rigorous grounds has not yet been accomplished. The situation is better for the iterative solution of the linear defect equations (6.3). Here, the multigrid method is used which inherits projection properties similar to those of the underlying finite element discretization. This can be exploited for deriving weighted

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

a posteriori error estimates which simultaneously control the discretization and the iteration error. The application of this concept has been described in Becker, Johnson and Rannacher (1995) for the Poisson model problem (3.2), in Becker (1998*a*) for the Stokes problem in primal-mixed formulation, and in Larson and Niklasson (1999) for semilinear elliptic problems. In both papers only energy-norm and L^2-norm error estimates have been given. However, the extension of this argument for also incorporating estimation of error functionals is straightforward.

Remark 6.2. We note that the evaluation of the *a posteriori* error estimate (6.4) involves only the solution of *linearized* problems. Hence, the whole error estimation may amount to only a relatively small fraction of the total cost for the solution process. This has to be compared to the usually much higher cost when working on non-optimized meshes.

6.2. Approximation of the 'exact' adjoint problem

The use of the error estimator $\eta_\omega(u_i)$ in (6.4) for mesh adaptation relies on the assumption that the remainder terms in the 'exact' error representations (2.18) or (2.27) are small compared to the residual terms and may therefore be neglected. This may even be justified for the simplest possible linearization using $\hat{u}_h = u_h$ and

$$J(u) - J(u_h) = \min_{\varphi_h \in V_h} \rho(u_h; z - \varphi_h) + R(\overline{uu_h}; e, e, z), \tag{6.5}$$

provided the mesh is sufficiently fine, depending on the stability properties of the continuous problem. However, there are several aspects which require special thought.

- The *a posteriori* error estimate (6.4) may lead to meshes that are rather coarse in areas of less importance for the computation of $J(u)$. This means that global error quantities like the energy-norm or the L^2-norm error may not be so small. In turn, the linearization may be less accurate globally.
- In situations when the problem is close to being singular (for example, close to a bifurcation point), the second derivative $A''(\overline{uu_h}; \cdot, \cdot, \cdot)$ may be very large, such that the remainder term may not be small enough to be neglected.

The estimation of the remainder term requires bounds for global integral expressions of the error $e = u - u_h$ in which the adjoint solution appears as a weight, for example $\|e^2 z\|_{L^1}$. It seems natural that the control of the linearization also reflects the sensitivity of the problem, that is, the dependence of the target quantity on the local residuals in terms of the adjoint solution z. Now, we want to discuss the possibility of making the

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

a posteriori error representation (6.5) reliable for estimating the error $J(u)-J(u_h)$ on meshes for which the remainder term cannot be guaranteed to be negligible. This must also include the generation of approximations $\hat{z}_h$ to the exact solution of the perturbed adjoint problem. Given a mesh $\mathbb{T}_h$, an associated finite element space V_h and an approximating solution $u_h \in V_h$, we have to approximate the solution $z \in V$ of the 'exact' adjoint problem (see Proposition 2.4)

$$A'(\overline{uu_h};\varphi,z) = J'(\overline{uu_h};\varphi), \quad \text{for all } \varphi \in V. \tag{6.6}$$

In Proposition 2.5, we have analysed the effect of linearizing (6.6) by replacing u by an improved approximation $\hat{u}_h \in \hat{V}_h$ in a richer finite element space $\hat{V}_h \supset V_h$:

$$A'(\overline{\hat{u}_h u_h};\varphi,z) = J'(\overline{\hat{u}_h u_h};\varphi), \quad \text{for all } \varphi \in V. \tag{6.7}$$

Then, the most straightforward approximation would be to solve this linearized adjoint problem on the same super-mesh for $\hat{z}_h \in \hat{V}_h$:

$$A'(\overline{\hat{u}_h u_h};\hat{\varphi}_h,\hat{z}_h) = J'(\overline{\hat{u}_h u_h};\hat{\varphi}_h), \quad \text{for all } \hat{\varphi}_h \in \hat{V}_h. \tag{6.8}$$

Of course, for solving (6.7), one could also use an even richer space than $\hat{V}_h$, depending on the particular properties of the adjoint problem. This variant will not be further discussed here. The resulting *a posteriori* error estimators may then be evaluated by one of the techniques described in Section 5. Involving more and more accurate approximations $\hat{u}_h$ and $\hat{z}_h$, the resulting error estimator on the current space V_h tends to become sharp. This approximation process can be realized in form of a feedback algorithm. For sequences of increasingly enriched spaces $V_i := V_{h_i}$ and $\hat{V}_{i+j} := \hat{V}_{h_{i+j}} \supset V_i$ $(i,j = 0,1,\dots)$, we write $u_i := u_{h_i}$, $z_i := z_{h_i}$, and analogously $\hat{u}_{i+j} := \hat{u}_{h_{i+j}}$, $\hat{z}_{i+j} := \hat{z}_{h_{i+j}}$. The resulting approximate error estimators are accordingly denoted by $\hat{\eta}^j_\omega(u_i)$.

Adaptive algorithm controlling linearization and discretization.

(0) Choose appropriate parameters $\varepsilon > 0$ and $0 < \varepsilon_0 \le \gamma\varepsilon$. Then, the algorithm aims at achieving effectivity index $I_{\text{eff}} = 1+\varepsilon$, while $\gamma \ll 1$ is a safety factor for the stopping criterion.

(1) Start from coarse meshes $\mathbb{T}_0$ and $\hat{\mathbb{T}}_0 := \mathbb{T}_0$ and compute the corresponding solutions u_0 and $\hat{z}_0$. Evaluate the approximate error estimator $\eta^0_\omega(u_0)$ and construct the new mesh $\mathbb{T}_1$.

(2) For $i \ge 1$, compute $u_i \in V_i$. Set $j = 0$.

(3) For $j \ge 0$, solve the adjoint problem with the bilinear form

$$A'(\overline{\hat{u}_{i+j}u_i};\cdot,\cdot)$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

on the mesh $\hat{\mathbb{T}}_{i+j}$ and evaluate the resulting error bound $\hat{\eta}^j_\omega(u_i)$. If

$$\left|\eta^j_\omega(u_i) - \eta^{j-1}_\omega(u_i)\right| > \varepsilon\, \eta^{j-1}_\omega(u_i),$$

then set $j := j+1$ and proceed.

(4) For $i \geq 0$, if $|\hat{\eta}^0_\omega(u_i) - \hat{\eta}^j_\omega(u_i)| \leq \varepsilon_0\, \hat{\eta}^j_\omega(u_i)$, then neglect the effect of linearization on the error estimator and continue with the mesh adaptation process using the error estimators $\eta_\omega(u_i) := \hat{\eta}^0_\omega(u_i)$. Further, if $\eta_\omega(u_i) \leq \mathrm{TOL}$ or $N_i \geq N_{\max}$, then stop; otherwise go to (3).

6.2.1. A semilinear example

We consider a semilinear problem of the form

$$L(u) := -\Delta u + s(u) = f \ \text{ in } \Omega, \quad u = 0 \ \text{ on } \partial\Omega, \tag{6.9}$$

defined on a bounded domain $\Omega \subset \mathbb{R}^2$. The nonlinearity $s(u)$ is assumed to have at most polynomial growth. Then, the corresponding semilinear form

$$A(u;\varphi) := (\nabla u, \nabla\varphi) + (s(u), \varphi)$$

is considered on the function space $V := H^1_0(\Omega)$. Its derivatives are

$$\begin{aligned} A'(u;\psi,\varphi) &= (\nabla\psi, \nabla\varphi) + (s'(u)\psi, \varphi) \\ A''(u;\xi,\psi,\varphi) &= (s''(u)\xi\psi, \varphi). \end{aligned}$$

The variational formulation of (6.9) seeks $u \in V$ such that

$$A(u;\varphi) = (f,\varphi), \quad \text{for all } \varphi \in V. \tag{6.10}$$

The Galerkin approximation of (6.9) uses finite element spaces $V_h \subset V$ as defined in Section 3. The discrete solutions $u_h \in V_h$ are determined by

$$A(u_h;\varphi_h) = (f,\varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{6.11}$$

In this case the general *a posteriori* error representation (2.27) results in the error estimate

$$|J(u) - J(u_h)| \leq \sum_{K\in\mathbb{T}_h} \rho_K\, \omega_K + |R(\overline{u\hat{u}_h u_h}; \hat{e}, e, z)|, \tag{6.12}$$

where the weights ω_K have the same form as in Proposition 3.1 and the residual terms ρ_K are defined by

$$\rho_K := \|R(u_h)\|_K + \|r(u_h)\|_{\partial K}$$

with $R(u_h)_{|K} := f + \Delta u_h - s(u_h)$, and $r(u_h)$ as defined above. The remainder term can be bounded by

$$|R(\overline{u\hat{u}_h u_h}; \hat{e}, e, z)| \leq \tfrac{1}{2}\max_{v\in\overline{u\hat{u}_h u_h}} \|s''(v)\|_{L^\infty} \|\hat{e}ez\|_{L^1}. \tag{6.13}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Numerical test (*from Vexler* (2000))
To illustrate the adaptive algorithm described above, we consider the nonlinear model problem

$$-\Delta(u-\hat{u}) + \lambda\big((u-\hat{u})^3 - (u-\hat{u})\big) = 0, \ \text{ in } \Omega, \quad u = 0, \ \text{ on } \partial\Omega, \qquad (6.14)$$

on the two-dimensional domain $\Omega := (0,1)^2$, with parameter $\lambda \geq 0$. The equilibrium solution is given by $\hat{u}(x_1,x_2) = 256x_1(1-x_1)x_2(1-x_2)$. Bifurcation from $\hat{u}$ occurs at the first critical value $\lambda_1^* = 2\pi^2 = 19.7392\ldots$. We want to compute the function value $u(0.75,0.75)$ with a reliable error bound, for the only slightly subcritical value $\lambda = 19$. We will see that by the above algorithm on a fixed mesh the effectivity index of the approximate error estimator can actually be driven to $I_{\text{eff}} = 1$. However, in the present almost singular situation, this requires an unacceptable amount of additional computational work on finer meshes. Therefore, we restrict our goal to achieving an effectivity of reasonable size $I_{\text{eff}} \leq 2$. The stopping parameters are chosen as $\varepsilon = 0.6$, $\varepsilon_0 = 0.06$. For this example, the described algorithm for controlling the linearization and discretization in the adjoint problem seems to work quite well. It achieves acceptable effectivity and the stopping criterion for the inner loops limits the amount of additional work.

Table 6.1. Control of linearization: numbers of cells N of the primary mesh $\mathbb{T}_i$ and $\hat{N}$ of the secondary mesh $\hat{\mathbb{T}}_i$

N	9	14	41	119	393	1367	5233	20338
$\hat{N}$	322	2256	1670	2247	2184	1367	5233	20338
I_{eff}	0.585	1.354	1.493	1.554	1.577	1.798	1.421	1.369

7. Realization for time-dependent problems

In the preceding sections, we have discussed the DWR method and some of its practical aspects for simple stationary diffusion and transport model problems. Now we want to extend this concept of adaptivity to nonstationary problems including parabolic as well as hyperbolic cases. We will see that in all these situations the general pattern of the DWR method remains the same: *variational formulation*; *error representation via duality argument*; *Galerkin orthogonality*; *approximate computation of adjoint solution*. This also applies to time-dependent problems provided that the discretization is organized as a Galerkin method in space and time. The following examples are used to discuss some aspects of the DWR method which are specific to the different types of differential equations. For simplicity, we restrict the

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

analysis here to *linear* problems. The treatment of *nonlinear time-dependent* problems by the DWR method is the subject of ongoing research.

7.1. Parabolic model problem: heat equation

We consider the heat-conduction problem

$$\begin{aligned} \partial_t u - \nabla\cdot(a\nabla u) &= f \qquad \text{in } Q_T, \\ u_{|t=0} &= u^0 \qquad \text{on } \Omega, \\ u_{|\partial\Omega} &= 0 \qquad \text{on } I, \end{aligned} \tag{7.1}$$

on a space-time region $Q_T := \Omega \times I$, where $\Omega \subset \mathbb{R}^d$, $d \geq 1$, and $I = (0,T)$; the coefficient a, the source term f and the initial value u^0 are assumed to be smooth. This model is used to describe diffusive transport of energy or certain species concentrations. The corresponding strong solution is unique in the space

$$V := H^1(0,T;L^2(\Omega)) \cap L^2(0,T;H_0^1(\Omega)).$$

The discretization of problem (7.1) is by a Galerkin finite element discretization simultaneously in space and time. We split the time interval $(0,T)$ into subintervals $I_n = (t_{n-1}, t_n]$, where

$$0 = t_0 < \cdots < t_n < \cdots < t_N = T, \qquad k_n := t_n - t_{n-1}.$$

At each time level t_n, let $\mathbb{T}_h^n$ be a regular finite element mesh as defined above with local mesh width $h_K = \operatorname{diam}(K)$, and let $V_h^n \subset H_0^1(\Omega)$ be the corresponding finite element subspace with d-linear shape functions. Extending the spatial mesh to the corresponding space-time slab $\Omega \times I_n$, we obtain a global space-time mesh consisting of $(d+1)$-dimensional cubes $Q_K^n := K \times I_n$. On this mesh, we define the global finite element space

$$\begin{aligned} V_{h,k} := \big\{ & v \in L^\infty(0,T;H_0^1(\Omega)) : \\ & v(\cdot,t)_{|Q_K^n} \in \tilde{Q}_1(K),\ v(x,\cdot)_{|Q_K^n} \in P_r(I_n) \text{ for all } Q_K^n \big\}, \end{aligned}$$

with some $r \geq 0$. For this setting, we introduce the space $\hat{V} := V \oplus V_{h,k}$. For functions from this space, we use the notation

$$v^{n+} := \lim_{t \to t_n + 0} v(t), \quad v^{n-} := \lim_{t \to t_n - 0} v(t), \quad [v]^n := v^{n+} - v^{n-}.$$

The discretization of problem (7.1) is based on a variational formulation which allows the use of discontinuous functions in time. This method, called the 'dG(r) method' (*discontinuous* Galerkin method in time), determines approximations $u_h \in V_{h,k}$ by

$$A_k(u_h, \varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h \in V_{h,k}, \tag{7.2}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

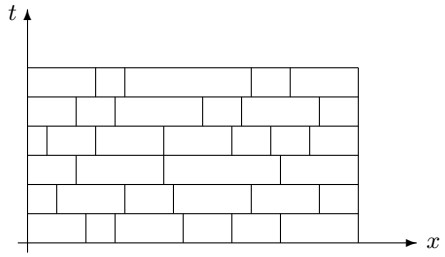


Fig. 7.1. Sketch of a space-time mesh with hanging nodes

with the bilinear form

$$A_k(v,\varphi) := \sum_{n=1}^{N} \int_{I_n} \big\{ (\partial_t v, \varphi) + (a\nabla v, \nabla\varphi) \big\}\, \mathrm{d}t + \sum_{n=2}^{N} ([v]^{n-1}, \varphi^{(n-1)+}) + (v^{0+}\varphi^{0+}),$$

and the linear functional

$$F(\varphi) := (f,\varphi) + (u^0, \varphi^{0+}),$$

which are both well defined on $\hat{V}$. We note that the continuous solution u also satisfies equation (7.2), which again implies Galerkin orthogonality for the error $e := u - u_h \in \hat{V}$ with respect to the bilinear form $A(\cdot,\cdot)$. Since the test functions $\varphi_h \in V_{h,k}$ may be discontinuous at times t_n, the global system (7.2) decouples and can be written in the form of a time-stepping scheme,

$$\int_{I_n} \big\{ (\partial_t u_h, \varphi_h) + (a\nabla u_h, \nabla\varphi_h) \big\}\, \mathrm{d}t + \big([u_h]^{n-1}, \varphi_h^{(n-1)+} \big) = \int_{I_n} (f, \varphi_h)\, \mathrm{d}t,$$

for all $\varphi_h \in V_h^n$, $n = 1, \ldots, N$. In the following, we consider for simplicity only the lowest-order case $r = 0$, the 'dG(0) method', which is equivalent to a variant of the backward Euler scheme. We concentrate on control of the spatial L^2-error $\|e^{N-}\|$ at the end-time $T = t_N$. To this end, we use the adjoint problem

$$\begin{aligned} &\partial_t z - a\Delta z = 0 \quad \text{in } \Omega \times I, \\ &z_{|t=T} = \|e^{N-}\|^{-1} e^{N-} \ \text{ in } \Omega, \ \ z_{|\partial\Omega} = 0 \ \text{ on } I, \end{aligned} \tag{7.3}$$

the strong solution of which automatically satisfies

$$A_k(\varphi, z) = J(\varphi) := \|e^{N-}\|^{-1} (\varphi^{N-}, e^{N-}), \quad \text{for all } \varphi \in \hat{V}.$$

Notice that, in this special situation, the adjoint solution z does not depend on the parameter k. Then, from the general results of Proposition 2.6, we

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

obtain the error representation

$$J(e) \;=\; F(z-\varphi_h) - A_k(u_h, z-\varphi_h), \tag{7.4}$$

for any $\varphi_h \in V_{h,k}$. By elementary reordering of terms in (7.4), we infer the estimate

$$\begin{aligned}\|e^{N-}\| \le \sum_{n=1}^{N} \sum_{K\in\mathbb{T}_h^n} \big|&(R(u_h), z-\varphi_h)_{K\times I_n} - (r(u_h), z-\varphi_h)_{\partial K\times I_n} \\ &- ([u_h]^{n-1}, (z-\varphi_h)^{(n-1)+})_K\big|,\end{aligned} \tag{7.5}$$

with the cell and edge residuals

$$\begin{aligned} R(u_h)_{|K\times I_m} &:= f + \nabla\cdot(a\nabla u_h) - \partial_t u_h, \\ r(u_h)_{|\Gamma\times I_m} &:= \begin{cases} \frac{1}{2} n\cdot[\nabla u_h], & \text{if } \Gamma \subset \partial K\backslash\partial\Omega, \\ 0, & \text{if } \Gamma \subset \partial\Omega, \end{cases} \end{aligned}$$

and an arbitrary approximation $\varphi_h \in V_{h,k}$. From this, we infer the following result.

Proposition 7.1. (Hartmann 1998) For the dG(0) finite element method applied to the heat equation (7.1), we have the *a posteriori* error estimate

$$\|e^{N+}\| \;\le\; \eta_\omega(u_h) := \sum_{n=1}^{N} \sum_{K\in\mathbb{T}_h^n} \big\{\rho_K^{n,1}\,\omega_K^{n,1} + \rho_K^{n,2}\,\omega_K^{n,2}\big\}, \tag{7.6}$$

where the cell-wise residuals and weights are defined by

$$\begin{aligned} \rho_K^{n,1} &:= \|R(u_h)\|_{K\times I_n} + h_K^{-1/2}\|r(u_h)\|_{\partial K\times I_n}, \\ \omega_K^{n,1} &:= \|z-\varphi_h\|_{K\times I_n} + h_K^{1/2}\|z-\varphi_h\|_{\partial K\times I_n}, \\ \rho_K^{n,2} &:= k_n^{-1/2}\|[u_h]^{n-1}\|_K, \qquad \omega_K^{n,2} := k_n^{1/2}\|(z-\varphi_h)^{(n-1)+}\|_K, \end{aligned}$$

with a suitable approximation $\varphi_h \in V_{h,k}$ to the adjoint solution z.

Remark 7.1. The statement of Proposition 7.1 extends to higher-order dG(r) methods and also to the related cG(r) methods (*continuous* Galerkin methods). In this case the flexibility in choosing the approximation φ_h in (7.5) can be used to replace the residual terms by

$$\rho_K^{n,1} := \|R(u_h) - \overline{R(u_h)}\|_{K\times I_n} + h_K^{-1/2}\|r(u_h) - \overline{r(u_h)}\|_{\partial K\times I_n},$$

for certain time-averages indicated by over-lines.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

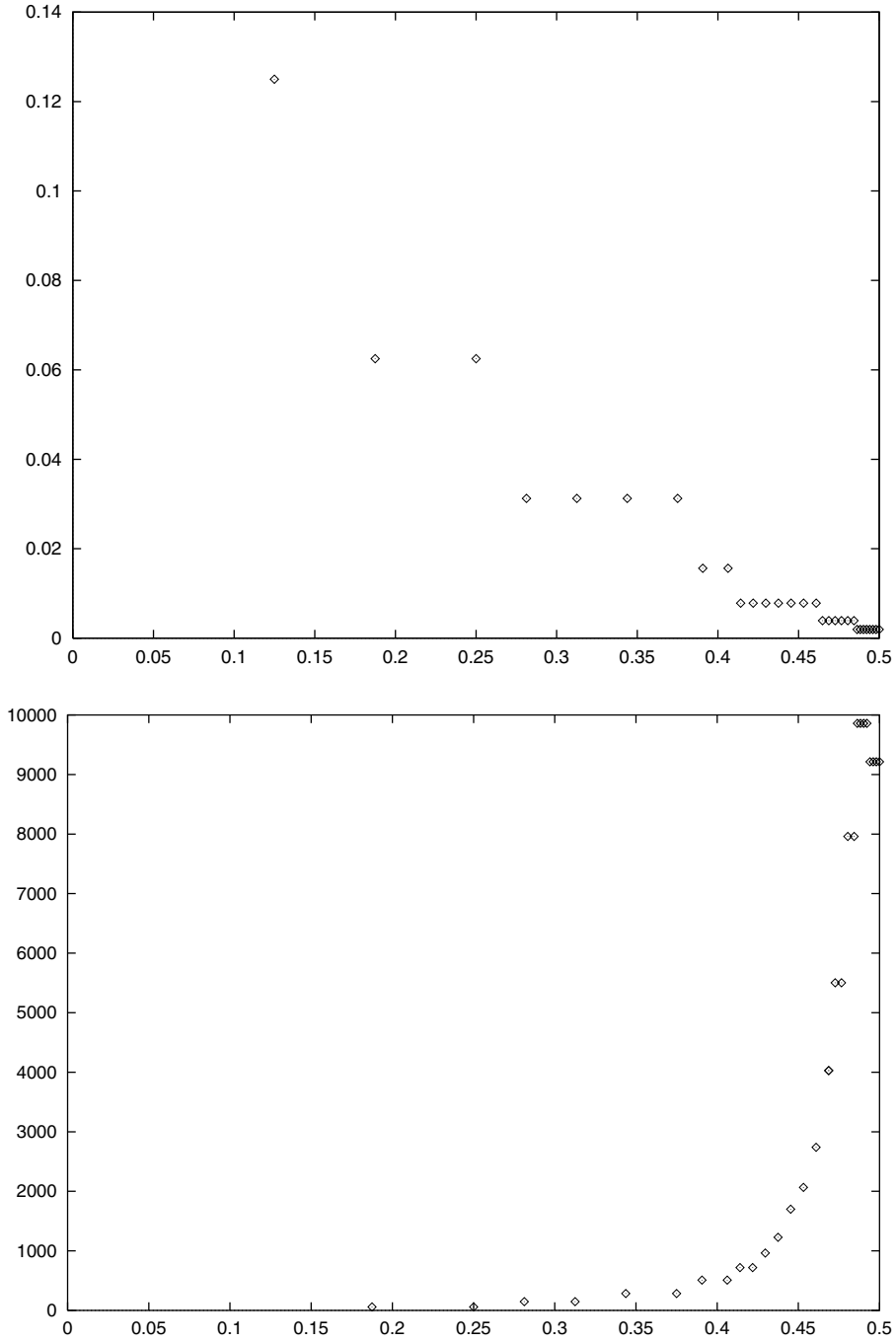


Fig. 7.2. Development of the time-step size (above) and the number N of mesh cells (below) by the DWR method over the time interval $I = [0, 0.5]$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Numerical test (*from Hartmann* (1998))
The performance of the error estimator (7.6) is illustrated by a simple test in two space dimensions where the (known) exact solution represents a smooth rotating bump with a suitably adjusted force f on the unit square. The computation is carried over the time interval $0 \leq t \leq T = 0.5$. In Figure 7.2, we show the development of the time-step size k and the spatial mesh complexity N, while Table 7.1 contains the corresponding results. Figure 7.3 shows a sequence of adapted meshes at successive times obtained by controlling the spatial L^2 error at the end-time $t_N = 0.5$. We clearly see the localizing effect of the weights in the error estimator, which suppress the influence of the residuals during the initial period.

Table 7.1. Results by simultaneous adaptation of spatial and time discretization by the DWR method ($M = \#$ time-steps, $N = \#$ mesh-cells)

M	N	$J(e)$	$\eta_\omega(u_h)$	I_{eff}
8	256	4.19e-2	3.04e-1	7.27
21	256	1.20e-2	7.14e-2	5.59
46	760	2.91e-3	6.60e-3	2.27
81	3472	7.92e-4	1.08e-3	1.36
119	9919	3.99e-4	6.64e-4	1.67

Notes and references
The traditional approach to *a posteriori* error estimation for parabolic problems combines 'energy norm' error estimates for spatial discretization (like the ones described in Section 4) with heuristic truncation error estimation (see, *e.g.*, Ziukas and Wiberg (1998) and Lang (1999)). The resulting error *indicators*, on principle, cannot reflect the true behaviour of the global error in time since the possible time-accumulation of the local truncation errors is not taken into account. Space-time duality arguments for error estimation in dG-finite element approximation of parabolic problems have been intensively used in a sequence of papers by Eriksson, Johnson and Larsson (Eriksson and Johnson 1991, 1995*a*, 1995*b*, 1995*c*, Eriksson *et al.* 1998) and by Estep and Larsson (1993). The same approach can also be used in the context of ordinary differential equations (see Estep (1995), Estep and French (1994) and Böttcher and Rannacher (1996)).

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

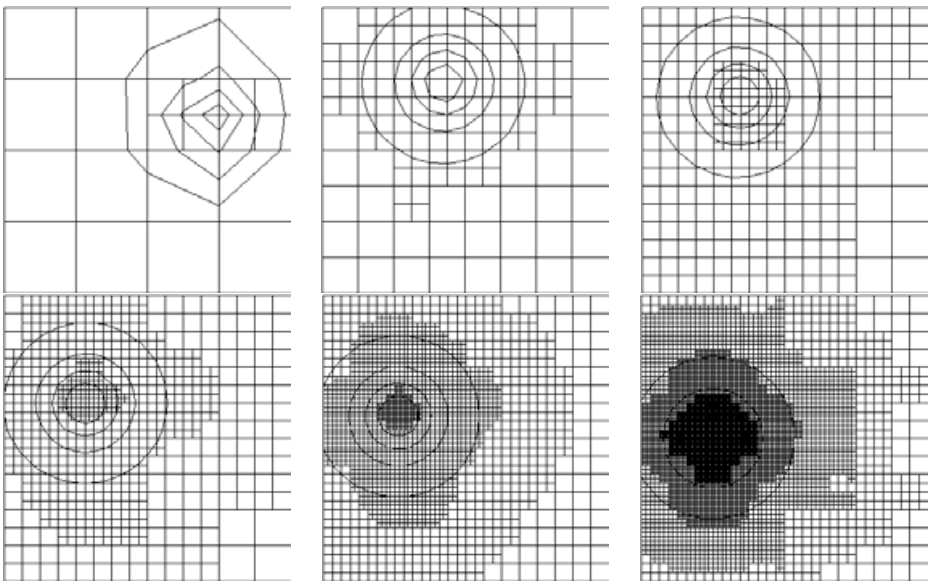


Fig. 7.3. Sequence of refined meshes for controlling the end-time error $\|e^{N-}\|$ shown at six consecutive times $t_n = 0.125, \ldots, 0.5$

7.2. Hyperbolic model problem: acoustic wave equation

We consider the acoustic wave equation

$$\begin{aligned} \partial_t^2 w - \nabla\cdot\{a\nabla w\} &= 0 \qquad \text{in } Q_T, \\ w_{|t=0} = w^0, \quad \partial_t w_{|t=0} &= v^0 \qquad \text{on } \Omega, \\ n\cdot a\nabla w_{|\partial\Omega} &= 0 \qquad \text{on } I, \end{aligned} \tag{7.7}$$

on a space-time region $Q_T := \Omega \times I$, where $\Omega \subset \mathbb{R}^d$, $d \geq 1$, and $I = (0, T)$; the elastic coefficient a may vary in space. This equation frequently occurs in the simulation of acoustic waves in gaseous or fluid media, seismics, and electrodynamics. We approximate problem (7.7) by a 'velocity-displacement' formulation which is obtained by introducing a new velocity variable $v := \partial_t w$. Then, the pair $u = \{w, v\}$ satisfies the system of equations

$$\partial_t w - v = 0, \quad \partial_t v - \nabla\cdot\{a\nabla w\} = 0, \tag{7.8}$$

with the natural solution space

$$V := [H^1(0, T; L^2(\Omega)) \cap L^2(0, T; H^1(\Omega))] \times H^1(0, T; L^2(\Omega)).$$

We derive a Galerkin discretization in space-time of (7.7) based on the formulation (7.8). To this end, the time interval $I = (0, T)$ is again decomposed into subintervals $I_n = (t_{n-1}, t_n]$ with length $k_n = t_n - t_{n-1}$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

and at each time level t_n a quadrilateral mesh $\mathbb{T}_h^n$ on Ω is chosen. On each time slab $Q^n := \Omega \times I_n$, we define intermediate meshes $\bar{\mathbb{T}}_h^n$ which are composed of the mutually finest cells of the neighbouring meshes $\mathbb{T}_h^{n-1}$ and $\mathbb{T}_h^n$, and obtain a decomposition of the time slab into space-time cubes $Q_K^n = K \times I_n, K \in \bar{\mathbb{T}}_h^n$. This construction is used in order to allow continuity in time of the trial functions when the meshes change with time. The discrete 'trial spaces' $V_{h,k}$ in space-time domain consist of functions that are $(d+1)$-linear on each space-time cell Q_K^n and globally *continuous* on Q_T. This prescription requires the use of 'hanging nodes' if the spatial mesh changes across a time level t_n. The corresponding discrete 'test spaces' $W_{h,k}$ consist of functions that are constant in time on each cell Q_K^n, while they are d-linear in space and globally continuous on Ω. On these spaces, we introduce the bilinear form

$$\begin{aligned} A(u,\varphi) := (\partial_t w, \xi)_{Q_T} - (v, \xi)_{Q_T} + (w(0), \xi(0)) \\ + (\partial_t v, \psi)_{Q_T} + (a\nabla w, \nabla\psi)_{Q_T} + (v(0), \psi(0)), \end{aligned}$$

and the linear functional

$$F(\varphi) = (w^0, \xi(0)) + (v^0, \psi(0)).$$

The Galerkin approximation of (7.7) seeks $u_h = \{w_h, v_h\} \in V_{h,k}$ satisfying

$$A(u_h, \varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h = \{\psi_h, \xi_h\} \in W_{h,k}. \tag{7.9}$$

For more details, we refer to Bangerth (1998) and Bangerth and Rannacher (1999). The scheme (7.9) is a 'Petrov–Galerkin' method. Since the solution $u = \{w, v\}$ also satisfies (7.9), we again have Galerkin orthogonality for the error $e := \{e^w, e^v\}$:

$$A(e, \varphi_h) = 0, \quad \varphi_h \in V_{h,k}. \tag{7.10}$$

This time-discretization scheme is called the 'cG(1) method' (*continuous* Galerkin method) in contrast to the dG method used in the preceding section. We note that, from this scheme, we can recover the standard Crank–Nicolson scheme in time (combined with a spatial finite element method):

$$\begin{aligned} (w^n - w^{n-1}, \varphi) - \tfrac{1}{2}k_n(v^n + v^{n-1}, \varphi) = 0, \\ (v^n - v^{n-1}, \psi) + \tfrac{1}{2}k_n(a\nabla(w^n + w^{n-1}), \nabla\psi) = 0. \end{aligned} \tag{7.11}$$

The system (7.11) splits into two equations, a discrete Helmholtz equation and a discrete L^2-projection.

In order to embed the present situation into the general framework laid out in Section 2, we introduce the spaces

$$\hat{V} := V \oplus V_{h,k}, \quad \hat{W} := V \oplus W_{h,k}.$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

We want to control the error in terms of a functional of the form

$$J(e) := (j, e^w)_{Q_T},$$

with some density function $j(x,t)$. To this end, we again use a duality argument in space-time employing the time-reversed wave equation

$$\begin{aligned} \partial_t^2 z^w - \nabla\cdot\{a\nabla z^w\} &= j \quad \text{in } Q_T, \\ z^w{}_{|t=T} = 0, \quad -\partial_t z^w{}_{|t=T} &= 0 \quad \text{on } \Omega, \\ n\cdot a\nabla z^w{}_{|\partial\Omega} &= 0 \quad \text{on } I. \end{aligned} \tag{7.12}$$

Its strong solution $z = \{-\partial_t z^w, z^w\} \in \hat{W}$ satisfies the variational equation

$$A(\varphi, z) = J(\varphi), \quad \text{for all } \varphi \in \hat{V}. \tag{7.13}$$

Then, from the general results of Proposition 2.3, we obtain the error identity

$$J(e) \;=\; F(z-\varphi_h) - A(w_h, z-\varphi_h), \tag{7.14}$$

for arbitrary $\varphi_h = \{\varphi_h^w, \varphi_h^v\} \in W_{h,k}$. Recalling the definition of the bilinear form $A(\cdot,\cdot)$, we obtain

$$\begin{aligned} |(j,w)_{Q_T}| \;\le\; \sum_{n=1}^{N} \sum_{K\in\mathbb{T}_h^n} \big| &(R^u(u_h), \partial_t z^w - \varphi_h^v)_{K\times I_n} \\ &- (R^v(u_h), z^w - \varphi_h^w)_{K\times I_n} - (r(u_h), z^w - \varphi_h^w)_{\partial K\times I_n}\big|, \end{aligned}$$

with the cell residuals

$$R^w(u_h)_{|K} := \partial_t w_h - v_h, \qquad R^v(u_h)_{|K} := \partial_t v_h - \nabla\cdot\{a\nabla w_h\},$$

and the edge residuals

$$r(w_h)_{|\Gamma\times I_m} := \begin{cases} \frac{1}{2} n\cdot[\nabla w_h], & \text{if } \Gamma \subset \partial K\backslash\partial\Omega, \\ 0, & \text{if } \Gamma \subset \partial\Omega. \end{cases}$$

From this, we infer the following result.

Proposition 7.2. (Bangerth 1998, Bangerth and Rannacher 1999) For the cG(1) finite element method applied to the acoustic wave equation (7.7), we have the *a posteriori* error estimate

$$|J(e^N)| \;\le\; \eta_\omega(u_h) := \sum_{n=1}^{N} \sum_{K\in\mathbb{T}_h^n} \big\{\rho_K^{n,1}\,\omega_K^{n,1} + \rho_K^{n,2}\,\omega_K^{n,2}\big\}, \tag{7.15}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

where the cell-wise residuals and weights are defined by

$$\rho_K^{n,1} := \|R_1(u_h)\|_{K\times I_n} + h_K^{-1/2}\|r(u_h)\|_{\partial K\times I_n},$$
$$\omega_K^{n,1} := \|\partial_t z^w - \varphi_h^v\|_{K\times I_n} + h_K^{1/2}\|z^w - \varphi_h^w\|_{\partial K\times I_n},$$
$$\rho_K^{n,2} := \|R_2(u_h)\|_{K\times I_n}, \qquad \omega_K^{n,2} := \|z^w - \varphi_h^w\|_{K\times I_n},$$

with a suitable approximation $\{\varphi_h^w, \varphi_h^v\} \in V_{h,k}$ to the adjoint solution $\{z^w, z^v\}$.

Below, we will compare the error estimator (7.15) with a simple heuristic 'energy error' indicator measuring the spatial smoothness of the computed solution w_h:

$$\eta_E(u_h) := \left(\sum_{n=1}^{N} \sum_{K\in\mathbb{T}_h^n} \rho_K^{n,3}(u_h)^2 \right)^{1/2}. \tag{7.16}$$

Numerical test (*from Bangerth* (1998))
The error estimator (7.15) is illustrated by a simple test: the propagation of an outward travelling wave on $\Omega = (-1,1)^2$ with a strongly heterogeneous coefficient. Layout of the domain and structure of the coefficient are shown in Figure 7.4. Boundary and initial conditions were chosen as follows:

$$n\cdot\{a\nabla u\} = 0 \quad \text{on } y = 1, \qquad w = 0 \quad \text{on } \partial\Omega\backslash\{y = 1\},$$
$$w_0 = 0, \quad v_0 = \theta(s - r)\,\exp\left(-|x|^2/s^2\right)\left(1 - |x|^2/s^2\right),$$

with $s = 0.02$ and $\theta(\cdot)$ the jump function. The region of origin of the wave field is significantly smaller than shown in Figure 7.4. Notice that the lowest

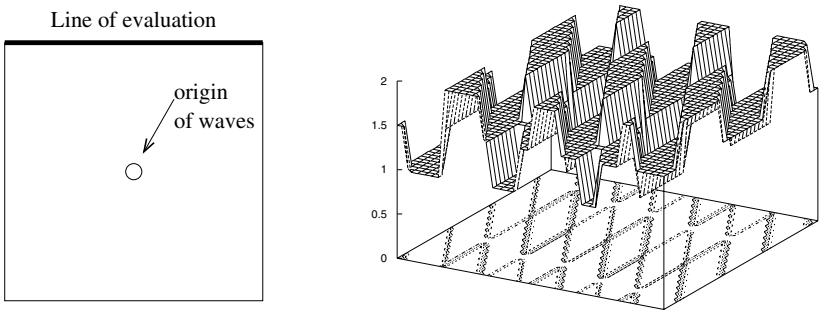


Fig. 7.4. Layout of the domain (left) and structure of the coefficient $a(x)$ (right)

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Table 7.2. Results obtained by adaptation of spatial discretization using the DWR method (reference value $J(w) \approx$ -4.515e-6, M =# time-steps, N =# mesh-cells)

Weighted estimator		Heuristic indicator	
$N \times M$	$J(w_h)$	$N \times M$	$J(w_h)$
327,789	-2.085e-6	327,789	-2.085e-6
920,380	-4.630e-6	920,380	-4.630e-6
2,403,759	-4.286e-6	2,403,759	-4.286e-6
1,918,696	-4.177e-6	5,640,223	-4.385e-6
2,975,119	-4.438e-6	10,189,837	-4.463e-6
6,203,497	-4.524e-6	17,912,981	-4.521e-6
		41,991,779	-4.517e-6

frequency in this initial wave field has wavelength $\lambda = 4s$; hence taking the common minimum ten grid points per wavelength would yield $62,500$ cells for the largest wavelength. Uniform grids quickly get to their limits in such cases. If we consider this example as a model of propagation of seismic waves in a faulted region of rock, then we would be interested in recording seismograms at the surface, here chosen as the top line Γ of the domain. A corresponding functional output is

$$J(w) = \int_0^T \int_\Gamma w(x,t)\ \omega(\xi,t)\,\mathrm{d}\xi\,\mathrm{d}t,$$

with a weight factor $\omega(\xi,t) = \sin(3\pi\xi)\sin(5\pi t/T)$, and end-time $T = 2$. The frequency of oscillation of this weight is chosen to match the frequencies in the wave field to obtain good resolution of changes.

In Figure 7.5, we show the grids resulting from refinement by the weighted error estimator (7.15) compared with the energy error indicator (7.16). Both resolve the wave field quite well, including reflections from discontinuities in the coefficient. The first additionally takes into account that the lower parts of the domain lie outside the domain of influence of the target functional if we truncate the time domain at $T = 2$; this domain of influence constricts to the top as we approach the final time, as is reflected by the produced grids. The meshes obtained in this way are obviously much more economical, without degrading the accuracy in approximating the quantity of interest (for more examples see Bangerth and Rannacher (1999)).

Remark 7.2. The evaluation of the *a posteriori* error estimate (7.15) requires a careful approximation of the adjoint solution z. Therefore we have used a higher-order method (bi-quadratic elements) for solving the space-

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

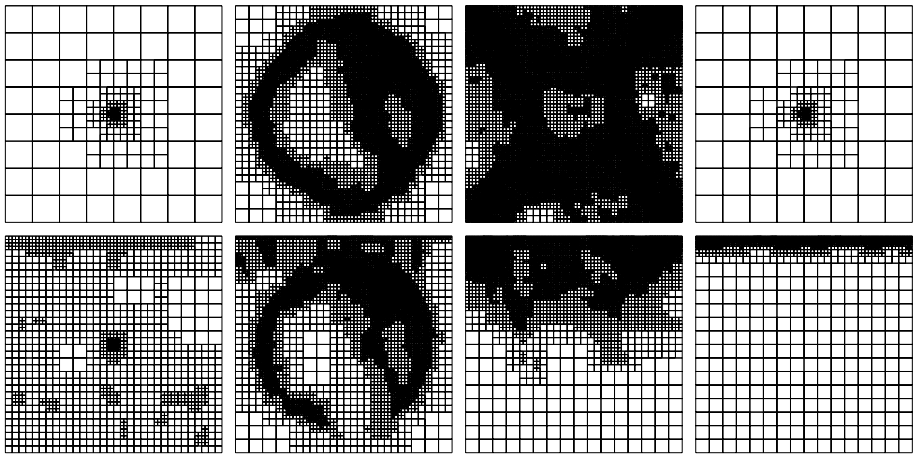


Fig. 7.5. Grids produced by the energy error indicator (top row) and by the dual-weighted estimator (bottom row) at times $t = 0, \frac{2}{3}, \frac{4}{3}, 2$

time adjoint problem, though this does not seem feasible for complex higher-dimensional problems.

Notes and references

The *a priori* error analysis of dG methods for the wave equation using space-time duality arguments has been initiated by Johnson (1993). Results obtained by heuristic ZZ-type indicators are discussed, for instance in Ziukas and Wiberg (1998). The extension of the adaptive cG method to the *elastic wave equation* is described by Bangerth (2000).

8. Application to incompressible viscous flow

In this section, we present an application of the DWR method to computing incompressible fluid flow governed by the classical Navier–Stokes equations. The purpose is to demonstrate that, by solving the global adjoint problem, one can actually capture the complex interaction of all physical mechanisms such that meshes are generated that are significantly more economical than those obtained by simple *ad hoc* adaptation. In this application (and in that presented in Section 9, below), we reach a point at which the concept of the DWR method laid out in Section 2 has to be taken only as a formal guideline for deriving meaningful *a posteriori* error estimates. A mathematically fully rigorous argument is not possible because of the complexity of the problem and the discretization procedure.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

8.1. The incompressible Navier–Stokes equations

We consider viscous incompressible Newtonian fluid flow modelled by the Navier–Stokes equations

$$-\nu\Delta v + v\cdot\nabla v + \nabla p = f, \qquad \nabla\cdot v = 0, \tag{8.1}$$

for the velocity v and the pressure p in a bounded domain $\Omega \subset \mathbb{R}^2$. Here, $\nu > 0$ is the normalized viscosity (density $\rho \equiv 1$), and the volume force is assumed as $f \equiv 0$. At the boundary $\partial\Omega$, the usual no-slip condition is imposed along rigid walls together with suitable inflow and 'free-stream' outflow conditions,

$$v|_{\Gamma_{\rm rigid}} = 0, \quad v|_{\Gamma_{\rm in}} = \hat{v}, \quad \nu\partial_n v - pn|_{\Gamma_{\rm out}} = 0.$$

As a concrete example, below, we will consider flow around a cylinder in a channel.

The variational formulation of (8.1) uses the function spaces

$$\hat{V} := L\times\hat{H}, \quad V := L\times H \subset \hat{V},$$

where $L := L^2(\Omega)$, $\hat{H} := H^1(\Omega)^2$, and $H := \{v\in H^1(\Omega)^2 : v_{|\Gamma_{\rm in}\cup\Gamma_{\rm rigid}} = 0\}$. For pairs $u = \{p, v\}$, $\varphi = \{q, w\} \in \hat{V}$, we define the semilinear form

$$A(u;\varphi) := \nu(\nabla v, \nabla w) + (v\cdot\nabla v, w) - (p, \nabla\cdot w) + (q, \nabla\cdot v)$$

and seek $u = \{p, v\} \in V + \{0, \hat{v}\}$, such that

$$A(u;\varphi) = (f, w), \quad \text{for all } \varphi = \{q, w\} \in V. \tag{8.2}$$

We assume that this problem possesses a (locally) unique solution that is stable, that is, the Fréchet derivative $A'(u;\cdot,\cdot)$ is coercive in the strong sense,

$$A'(u;\varphi,\varphi) \geq \gamma\|\varphi\|^2, \quad \varphi \in V.$$

For discretizing this problem, we use a finite element method based on the quadrilateral Q_1/Q_1-Stokes element with globally continuous (piecewise isoparametric) bilinear shape functions for both unknowns, pressure and velocity. As described before, we allow 'hanging' nodes while the corresponding unknowns are eliminated by linear interpolation. The corresponding finite element subspaces are denoted by

$$L_h \subset L, \quad \hat{H}_h \subset \hat{H}, \quad H_h \subset H, \quad \hat{V}_h := L_h\times\hat{H}_h, \quad V_h := L_h\times H_h,$$

and $\hat{v}_h \in \hat{H}_h$ is a suitable interpolation of the boundary function $\hat{v}$. This construction is oriented by the situation of a polygonal domain Ω for which the boundary $\partial\Omega$ is exactly matched by the mesh domain $\Omega_h := \cup\{K \in \mathbb{T}_h\}$. In the case of a general curved boundary (as in the flow example, below) some standard modifications are necessary, the description of which is omitted here.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

In order to obtain a stable discretization of (8.2) in these spaces with 'equal-order interpolation' of pressure and velocity, we use the least-squares technique proposed by Hughes, Franca and Balestra (1986). Following Hughes and Brooks (1982), a similar approach is employed for stabilizing the convective term. The Navier–Stokes system can be written in vector form for the unknown $u = \{p, v\} \in \hat{V}$ as

$$A(u) := \begin{bmatrix} -\nu\Delta v + v\cdot\nabla v + \nabla p \\ \nabla\cdot v \end{bmatrix} = \begin{bmatrix} f \\ 0 \end{bmatrix} =: F.$$

Then, the strong solution $u = \{p, v\}$ of (8.1) satisfies $A(u) = F$ in the L^2-sense. With the operator $A(\cdot)$, we associate a derivative $A'(u)$ at u and an approximation $S(u)$ which act on $\varphi = \{q, w\} \in \hat{V}$ via

$$A'(u)\varphi := \begin{bmatrix} -\nu\Delta w + v\cdot\nabla w + w\cdot\nabla v + \nabla q \\ \nabla\cdot w \end{bmatrix}, \qquad S(u)\varphi := \begin{bmatrix} v\cdot\nabla w + \nabla q \\ 0 \end{bmatrix}.$$

With this notation, we introduce the stabilized form

$$A_h(u;\varphi) := A(u;\varphi) + (A(u) - F, S(u)\varphi)_h,$$

with the mesh-dependent inner product and norm

$$(v, w)_h := \sum_{K\in\mathbb{T}_h} \delta_K\,(v, w)_K, \quad \|v\|_h = (v, v)_h^{1/2}.$$

The stabilization parameter is chosen according to

$$\delta_K = \alpha\big(\nu h_K^{-2}, \beta\,|v_h|_{K;\infty} h_K^{-1}\big)^{-1}, \quad \delta := \max_{K\in\mathbb{T}_h} \delta_K, \tag{8.3}$$

with the heuristic choice $\alpha = \frac{1}{12}$, $\beta = \frac{1}{6}$. Now, in the discrete problems, we seek $\{p_h, v_h\} \in V_h + \hat{u}_h$, such that

$$A_h(u_h;\varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h \in V_h. \tag{8.4}$$

This approximation is fully consistent in the sense that the solution $u = \{p, v\}$ also satisfies (8.4). This again implies Galerkin orthogonality for the error $e = \{e^p, e^v\} := \{p - p_h, v - v_h\}$ with respect to the form $A_h(\cdot;\cdot)$:

$$A_h(u;\varphi_h) - A_h(u_h;\varphi_h) = 0, \quad \varphi_h \in V_h. \tag{8.5}$$

We note that, instead of S, one can also use $-S^*$ in the stabilized form $A_h(\cdot;\cdot)$ (see the compressible flow example below).

We now turn to the question of *a posteriori* error estimation in the scheme (8.4). Let the goal of the computation again be the evaluation of a quantity $J(u)$ expressed in terms of a linear functional $J(\cdot)$ on $\hat{V}$ (for simplicity). Here, we think of local quantities such as point values of pressure, drag and lift coefficients or averages of vorticity localized to certain recirculation zones. In order to control the error $J(u)$–$J(u_h)$, we consider the h-dependent

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

adjoint problem

$$\tilde{A}_h'(u;\varphi,z) := A'(u;\varphi,z) + (S(u)\varphi, S(u)z)_h \;=\; J(\varphi), \quad \text{for all } \varphi \in \hat{V}, \tag{8.6}$$

with the approximate derivative $S(u)$ defined above. Since the stabilizing bilinear form $(S(u)\cdot, S(u)\cdot)_h$ is coercive, the strong coercivity of $A'(u;\cdot,\cdot)$ also implies the (unique) solvability of the adjoint problem (8.6). With this notation, we have the following result.

Proposition 8.1. For a (linear) functional $J(\cdot)$ let $z = \{z^p, z^v\} \in V$ be the solution of the linearized adjoint problem (8.6). Then we have the *a posteriori* error estimate

$$|J(e)| \;=\; \eta_\omega(u_h) \;:=\; \sum_{K\in\mathbb{T}_h} \left\{ \sum_{\alpha\in\{p,v\}} \rho_K^\alpha \omega_K^\alpha + \cdots \right\} + R_h, \tag{8.7}$$

where the residual terms and weights are given by

$$\rho_K^p := \|R^p(u_h)\|_K, \quad \omega_K^p := \|z^p - \varphi_h^p\|_K,$$
$$\rho_K^v := \|R^v(u_h)\|_K + h_K^{-1/2}\|r^v(u_h)\|_{\partial K},$$
$$\omega_K^v := \|z^v - \varphi_h^v\|_K + h_K^{1/2}\|z^v - \varphi_h^v\|_{\partial K} + \delta_K\|v_h\cdot\nabla(z^v - \varphi_h^v) + \nabla(z^p - \varphi_h^p)\|_K,$$

with suitable approximations $\{\varphi_h^p, \varphi_h^v\} \in V_h$ to $\{z^p, z^v\}$. The dots '$\ldots$' in (8.7) stand for additional terms measuring the errors in approximating the inflow data and the curved cylinder boundary. The remainder term can be bounded by

$$|R_h| \;\le\; \|e^v\|\,\|\nabla e^v\|\,\|z^v\|_\infty + \delta\, C(u,e,z), \tag{8.8}$$

with a function $C(u,e,z)$ linear in e. For simplicity, the explicit dependence of the stabilization parameter δ on the solution is neglected.

Proof. We apply the abstract result of Proposition 2.6. Neglecting the errors due to the approximation of the inflow boundary data $\hat{v}$ as well as the curved cylinder boundary S, we have to evaluate the residual $\rho(u_h;\varphi) := F(\varphi) - A_h(u_h,\varphi)$ for $\varphi = \{q, w\} \in V$. By definition,

$$\begin{aligned}\rho(u_h;\varphi) \;=\; & (f,w) - \nu(\nabla v_h, \nabla w) - (v_h\cdot\nabla v_h, w) + (p_h, \nabla\cdot w) \\ & - (q, \nabla\cdot u_h) - (A(u_h) - F, S(u_h)\varphi)_h.\end{aligned}$$

Splitting the integrals into their contributions from each cell $K \in \mathbb{T}_h$ and integrating cell-wise by parts yields, analogously to deriving (3.18),

$$\begin{aligned}\rho(u_h;\varphi) \;=\; \sum_{K\in\mathbb{T}_h} \big\{ & (R^v(u_h), w)_K + (r^v(u_h), w)_{\partial K} + (q, R^p(u_h))_K \\ & + \delta_K(R^v(u_h), v_h\cdot\nabla w + \nabla q)_K \big\}.\end{aligned}$$

From this, we obtain the error estimate (8.7) if we set $\varphi := z - z_h$. It remains

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

to estimate the remainder term R_h which, observing that $A(u)-F=0$, in the present situation has the form

$$R_h \;=\; (S(u)e, S(u)z)_h) - (A'(u)e, S(u)z)_h + \int_0^1 A_h''(u_h+se; e, e, z)s\,\mathrm{d}s.$$

We recall that, for arguments $u=\{v,p\}, \varphi=\{q,w\}$ and $z=\{z^p,z^v\}$,

$$A'(u)\varphi := \begin{bmatrix} -\nu\Delta w + v\cdot\nabla w + w\cdot\nabla v + \nabla q \\ \nabla\cdot w \end{bmatrix}, \qquad S(u)\varphi := \begin{bmatrix} v\cdot\nabla w + \nabla q \\ 0 \end{bmatrix},$$

$$A''(u)\varphi z := \begin{bmatrix} w\cdot\nabla z^v + z^v\cdot\nabla w \\ 0 \end{bmatrix}, \qquad S'(u)\varphi z := \begin{bmatrix} w\cdot\nabla z^v \\ 0 \end{bmatrix}.$$

Hence, the first two terms in the remainder R_h can be written in the form

$$(S(u)e, S(u)z)_h - (A'(u)e, S(u)z)_h \;=\; -(-\nu\Delta e^v + e^v\cdot\nabla v, v\cdot\nabla z^v + \nabla z^p)_h.$$

Further, the first derivative of $A_h(\cdot;\cdot)$ has, for the arguments $u=\{p,v\}$, $\varphi=\{q,w\}\in V$ and $z=\{z^p,z^v\}\in V$, the explicit form

$$A_h'(u;\varphi,z) \;=\; (A'(u)\varphi, z) + (A'(u)\varphi, S(u)z)_h + (A(u), S'(u)\varphi z)_h,$$

with $A'(u)\varphi$ and $S(u)z$ defined as above. Since $S''(u)\equiv 0$, the second derivative of $A_h(\cdot;\cdot)$ has, for arguments $u_s=\{p_s,v_s\}$, $e=\{e^p,e^v\}$ and $z=\{z^p,z^v\}$, the form

$$\begin{aligned} A_h''(u_s;e,e,z) \;=\; & 2(e^v\cdot\nabla e^v, z^v) + 2(e^v\cdot\nabla e^v, v_s\cdot\nabla z^v + \nabla z^p)_h \\ & + 2\,(-\nu\Delta e^v + v_s\cdot\nabla e^v + e^v\cdot\nabla v_s + \nabla e^p, e^v\cdot\nabla z^v)_h. \end{aligned}$$

From this, we easily infer the proposed bound for the remainder term. □

We will compare the weighted error estimator $\eta_\omega(u_h)$ of Proposition 8.1 with several heuristically based error indicators. In all cases the mesh refinement is driven by local 'error indicators' $\eta_K=\eta_K(v,p)$ which are cheaply obtained from the computed solution $\{p_h,v_h\}$. Examples of common *heuristic* indicators are:

Vorticity. $\eta_K := \|\nabla\times v_h\|_K$,

Pressure-gradient. $\eta_K := \|\nabla p_h\|_K$,

'Energy error'. $\eta_K := \|R^p(u_h)\|_K + \|R^v(u_h)\|_K + h_K^{1/2}\|r^v(u_h)\|_{\partial K}$, with the cell and edge residuals defined by

$$R^p(u_h)_{|K} := \nabla\cdot v_h,$$

$$R^v(u_h)_{|K} := -\nu\Delta v_h + v_h\cdot\nabla v_h + \nabla p_h,$$

$$r^v(u_h)_{|\Gamma} := \begin{cases} -\frac{1}{2}[\nu\partial_n v_h - p_h n], & \text{if } \Gamma\not\subset\partial\Omega, \\ -(\nu\partial_n v_h - p_h n), & \text{if } \Gamma\subset\Gamma_{\text{out}}, \\ 0, & \text{if } \Gamma\subset\Gamma_{\text{rigid}}\cup\Gamma_{\text{in}}. \end{cases}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

The vorticity and the pressure-gradient indicators measure the 'smoothness' of the computed solution $\{v_h, p_h\}$ while the 'energy error' indicator additionally contains information concerning local conservation of mass and momentum. However, neither of them contains information about the effect of changing the local mesh size on the error in the target quantity. This is only achieved by the 'weighted' indicators derived from $\eta_\omega(u_h)$.

Numerical results (*from Becker, Braack and Rannacher* (2000*a*))
We consider the flow around the cross section of a cylinder with surface S in a channel with a narrowed outlet (see Figure 8.1). Here, a quantity of physical interest is, for example, the 'drag coefficient' defined by

$$J(v,p) := \frac{2}{U^2 D} \int_S n \cdot \sigma(v,p)\psi \, \mathrm{d}s, \tag{8.9}$$

where $\psi := (0,1)^T$. Here, $\sigma(v,p) = \frac{1}{2}\nu(\nabla v + \nabla v^T) + pI$ is the stress acting on the cylinder, D is its diameter, and $U := \max|\hat{v}|$ is the reference inflow velocity. In this example, the Reynolds number is $\mathrm{Re} = UD/\nu = 50$, such that the flow is stationary.

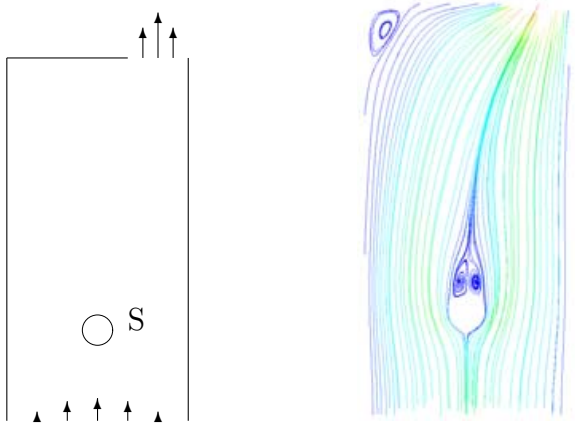


Fig. 8.1. Configuration and streamline plot of the 'flow around a cylinder'

In Figure 8.2, we compare the results for approximating the drag coefficient $J(v,p)$ on meshes which are constructed by using the different error indicators with that obtained on uniformly refined meshes. It turns out that all the above indicators perform equally weakly. A better result is obtained by using the 'weighted indicators' derived from (8.7). We clearly see the advantage of this error estimator which reflects the sensitivity of the error $J(v,p) - J(v_h,p_h)$ with respect to changes of the cell-wise residuals under mesh refinement.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

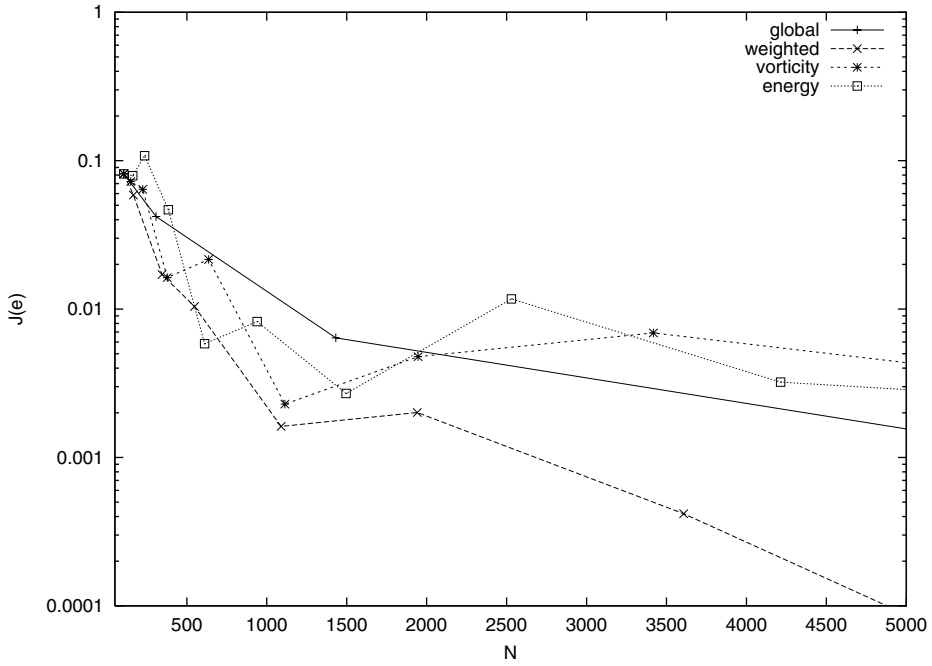


Fig. 8.2. Mesh efficiency obtained by global uniform refinement ('global' +), the weighted indicator ('weighted' ×), the vorticity indicator ('vorticity' ∗), and the energy indicator ('energy' □)

In the estimate (8.7) the additional terms representing the errors in approximating the inflow data and the curved boundary component S are neglected; they can be expected to be small compared to the other residual terms. The bounds for the dual solution $\{z^p, z^v\}$ are obtained computationally using patch-wise biquadratic interpolations as discussed in Section 5. This avoids the use of interpolation constants. The quantitative results of Figure 8.2 and the corresponding meshes shown in Figure 8.3 confirm the superiority of the weighted error indicator in computing local quantities.

Remark 8.1. In our computation the drag coefficient has actually been evaluated from the formula

$$J(v,p) := \frac{2}{\hat{H}^2 D} \int_\Omega \left\{ \sigma(p,v)\nabla\bar{\psi} + \nabla\cdot\sigma(p,v)\bar{\psi} \right\} \mathrm{d}x, \tag{8.10}$$

where $\bar{\psi}$ is an extension of the directional vector $\psi := (0,1)^T$ from S to Ω with support along S. By integration by parts, one sees that this definition is independent of the choice of $\bar{\psi}$ and that it is equivalent to the original one (8.9) as a contour integral. However, on the discrete level the two formulations differ. In fact, computational experience shows that formula (8.10)

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

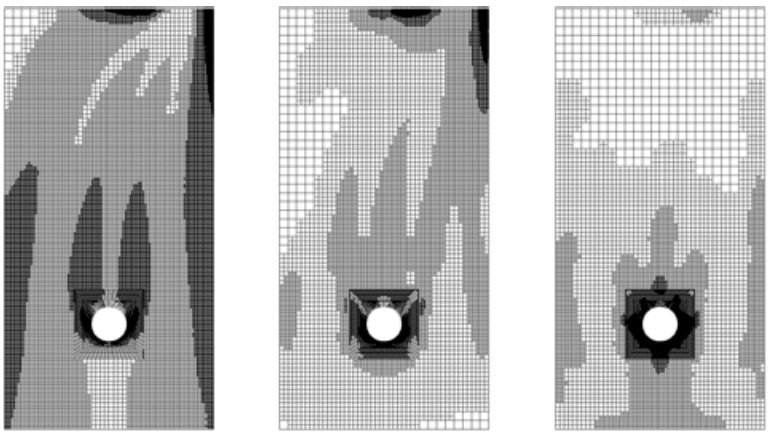


Fig. 8.3. Meshes with about 5000 cells obtained by the vorticity indicator (left), the 'energy' indicator (middle), and the weighted indicator (right)

yields significantly more accurate approximations of the drag coefficient (see Giles, Larsson, Levenstam and Süli (1997) and Becker (1998*b*)).

Notes and references

Energy norm-type *a posteriori* error estimates for finite element approximations of the Stokes and Navier–Stokes equations have been derived by Verfürth (1993), Bernardi, Bonnon, Langouët and Métivet (1995), Oden, Wu and Ainsworth (1993) and Ainsworth and Oden (1997*b*). An error analysis based on duality arguments for estimating the L^2 error was developed by Johnson and Rannacher (1994) and Johnson, Rannacher and Boman (1995). Another variant of this technique also including functional error estimation along the lines described in Section 4 can be found in Machiels *et al.* (1998) and Machiels, Peraire and Patera (1999). In Becker and Rannacher (1996*a*) and Becker (1995, 1998*b*), the application of the DWR approach to drag and lift computation was developed for the 'cylinder flow' benchmark described in Schäfer and Turek (1996).

9. Application to thermal and reactive flow

Now the complexity of the preceding example (incompressible Navier–Stokes equations) is further increased by including compressibility effects due to energy transfer and heat-release by chemical reactions. This constitutes the most complex situation the DWR method has been applied to yet. Though the model involves various interacting physical mechanisms (diffusion, transport and reaction) and several physical quantities of different sizes (density,

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

velocity, temperature, chemical species), the general approach works surprisingly well. It not only detects the quantitative dependencies of the target quantities on the local cell residuals but, by careful evaluation of the adjoint solution, also the interaction of the different solution components.

9.1. Low-Mach-number heat-driven flow

We concentrate on so-called 'low-Mach-number' gas flows where density variations are mainly due to temperature gradients. Such conditions often occur in chemically reactive flows and are characterized by hydrodynamically incompressible behaviour. Below, we will consider 'heat-driven' natural convection in a cavity as a typical example.

The underlying mathematical model is the full set of the (stationary) compressible Navier–Stokes equations in the so-called 'low-Mach-number approximation' due to the low speed of the resulting flow. Accordingly, the total pressure is split like $P(x) = P_{\rm th} + p(x)$ into a thermodynamic part $P_{\rm th}$, which is constant in space and used in the gas law, and a hydrodynamic part $p(x) \ll P_{\rm th}$ used in the momentum equation. Then, the governing system of conservation equations can be written in the following form:

$$\begin{aligned} \nabla\cdot v - T^{-1} v\cdot\nabla T &= 0, \\ \rho v\cdot\nabla v + \nabla\cdot\tau + \nabla p &= (\rho-\rho_0)\, g, \\ \rho v\cdot\nabla T - \nabla\cdot(\kappa\nabla T) &= f_T, \end{aligned} \tag{9.1}$$

supplemented by the law of an ideal gas $\rho = P_{\rm th}/RT$. The stress tensor is given by $\tau = -\mu\{\nabla v + (\nabla v)^T - \frac{2}{3}(\nabla\cdot v) I\}$.

In the model case considered below, there are no heat sources (*e.g.*, due to chemical reactions), that is, $f_T = 0$. The boundary conditions are no-slip for the velocity along the whole boundary, $v_{|\partial\Omega} = 0$, Neumann boundary conditions for the temperature along adiabatic walls, $\partial_n T_{|\Gamma_N} = 0$, and Dirichlet conditions for the temperature along the heated or cooled walls, $T_{|\Gamma_D} = \hat{T}$.

The variational formulation of (9.1) uses the following semilinear form defined for triples $u = \{p, v, T\}$, $\varphi = \{\xi, \psi, \theta\}$:

$$\begin{aligned} A(u;\varphi) \;:=\; & (\nabla\cdot v - T^{-1} v\cdot\nabla T, \xi) + (\rho v\cdot\nabla v, \psi) - (\tau, \nabla\psi) - (p, \nabla\cdot\psi) \\ & - (p, \nabla\cdot\psi) - (\rho g, \psi) + (\rho v\cdot\nabla T, \theta) + (\kappa\nabla T, \nabla\theta). \end{aligned}$$

Further, we define the functional

$$F(\varphi) \;:=\; -(\rho_0 g, \psi).$$

The natural solution spaces are

$$\hat{V} = L^2(\Omega)/\mathbb{R} \times H^1(\Omega)^2 \times H^1(\Omega), \quad V := L^2(\Omega)/\mathbb{R} \times H^1_0(\Omega)^2 \times H^1_0(\Gamma_D;\Omega),$$

where $H^1_0(\Gamma_D;\Omega) := \{\theta \in H^1(\Omega),\, \theta = 0 \text{ on } \Gamma_D\}$. With this notation, the

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

variational form of (9.1) seeks $u = \{p, v, T\} \in V + \hat{u}$, with $\hat{u} := \{0, 0, \hat{T}\}$, satisfying

$$A(u; \varphi) = F(\varphi), \quad \text{for all } \varphi \in V, \tag{9.2}$$

where ρ is considered as a (nonlinear) coefficient determined by the temperature through the equation of state $\rho = P_{\text{th}}/RT$. For more details on the derivation of this model, we refer to Braack and Rannacher (1999), and the literature cited therein. Here, we assume that (9.2) possesses a (unique) solution $u \in \hat{V}$, which is stable in the sense that the corresponding Fréchet derivative $A'(u; \cdot, \cdot)$ is strongly coercive.

The discretization of the system (9.2) again uses the continuous Q_1-finite element for all unknowns and employs least-squares stabilization for the velocity-pressure coupling as well as for the transport terms. We do not explicitly state the corresponding discrete equations since they have an analogous structure, as already seen in the preceding example of the incompressible Navier–Stokes equations. The derivation of the related adjoint problem and the resulting *a posteriori* error estimates follows the same line of argument. For economy reasons, we do not use the full Jacobian of the coupled system in setting up the adjoint problem, but only include its dominant parts. The same simplification is used in the nonlinear iteration process. For details, we refer to Braack and Rannacher (1999) and Becker, Braack and Rannacher (2000*a*). Below, we again use the mesh-dependent inner product and norm

$$(v, \psi)_h := \sum\nolimits_{K \in \mathbb{T}_h} \delta_K \, (v, \psi)_K, \quad \|v\|_h = (v, v)_h^{1/2},$$

with some stabilization parameters δ_K. The discrete problems seek $u_h = \{p_h, v_h, T_h\} \in V_h + \hat{u}_h$, satisfying

$$A_h(u_h; \varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h \in V_h, \tag{9.3}$$

with the stabilized form

$$A_h(u_h; \varphi_h) := A(u_h; \varphi_h) + (A(u_h) - F, S(u_h)\varphi_h)_h.$$

Here, the operator $A(u_h)$ is the generator of the form $A(u_h; \cdot)$, and the operator $S(u_h)$ in the stabilization term is chosen according to

$$S(u_h) := \begin{bmatrix} 0 & \text{div} & 0 \\ \nabla & \rho_h v_h \cdot \nabla + \nabla \cdot \mu \nabla & 0 \\ -T_h^{-1} v_h \cdot \nabla & 0 & \rho_h v_h \cdot \nabla + \nabla \cdot \kappa \nabla \end{bmatrix}.$$

As on the continuous level, the discrete density is determined by the temperature through the equation of state $\rho_h := P_{\text{th}}/RT_h$. We introduce the following notation for the cell residuals of the solution $u_h = \{p_h, v_h, T_h\}$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

of (9.3):

$$\begin{aligned} R^p(u_h)_{|K} &:= \nabla\cdot v_h - T_h^{-1} v_h\cdot\nabla T_h, \\ R^v(u_h)_{|K} &:= \rho_h v_h\cdot\nabla v_h - \nabla\cdot(\mu\nabla v_h) + \nabla p_h + (\rho_0-\rho_h)g, \\ R^T(u_h)_{|K} &:= \rho_h v_h\cdot\nabla T_h - \nabla\cdot(\kappa\nabla T_h) - f_T. \end{aligned}$$

Further, we define the edge residuals

$$r^v(u_h)_{|\Gamma} := \begin{cases} -\frac{1}{2}[\nu\partial_n v_h - p_h n], & \text{if } \Gamma \not\subset \partial\Omega, \\ -(\nu\partial_n v_h - p_h n), & \text{if } \Gamma \subset \Gamma_{\text{out}}, \\ 0, & \text{if } \Gamma \subset \Gamma_{\text{rigid}} \cup \Gamma_{\text{in}}. \end{cases}$$

$$r^T(u_h)_{|\Gamma} := \begin{cases} -\frac{1}{2}[\kappa\partial_n T_h], & \text{if } \Gamma \not\subset \partial\Omega, \\ -\kappa\partial_n T_h, & \text{if } \Gamma \subset \Gamma_N, \\ 0, & \text{if } \Gamma \subset \Gamma_D. \end{cases}$$

with $[\cdot]$ again denoting the jump across an interior edge Γ. These quantities will be needed below for defining the *a posteriori* error estimator. Using this notation the stabilizing part in $A_h(\cdot;\cdot)$ can be written in the form

$$\begin{aligned} (A(u_h), S(u_h)\varphi)_h &= (R^p(u_h), \nabla\cdot\psi)_h \\ &\quad + (R^v(u_h), \nabla\xi + \rho_h v_h\cdot\nabla\psi + \nabla\cdot(\mu\nabla\psi))_h \\ &\quad + (R^T(u_h), \rho_h v_h\cdot\nabla\theta + \nabla\cdot(\kappa\nabla\theta) - T_h^{-1} v_h\cdot\nabla\xi)_h, \end{aligned}$$

for $\varphi = \{\xi, \psi, \theta\}$. These terms comprise stabilization of the stiff velocity-pressure coupling in the low-Mach-number case and stabilization of transport in the momentum and energy equation case as well as the enforcement of mass conservation. The parameters $\delta_K = \{\delta_K^p, \delta_K^v, \delta_K^T\}$ may be chosen differently in the three equations following rules like (8.3). The stability of this discretization has been investigated in Braack (1998).

Now, we turn to the question of *a posteriori* error estimation in the scheme (9.3). Let $J(\cdot)$ again be a (for simplicity) linear functional defined on $\hat{V}$ for evaluating the error $e = \{e^p, e^v, e^T\}$. As for the previous section on incompressible flow, the adjoint problem is again set up using a reduced Jacobian in order to guarantee existence of the adjoint solution. Accordingly, we consider the h-dependent adjoint problem

$$\tilde{A}_h'(u;\varphi,z) := A'(u;\varphi,z) + (S(u)\varphi, S(u)z)_h = J(\varphi), \quad \text{for all } \varphi \in \hat{V}. \tag{9.4}$$

Then, from Proposition 2.6, we obtain for the present situation the following.

Proposition 9.1. Let $z = \{r, w, S\} \in V$ be the solution of the linearized adjoint problem (9.4). Then, we have the *a posteriori* error estimate

$$|J(e)| \approx \eta_\omega(u_h) := \sum_{K\in\mathbb{T}_h} \left\{ \sum_{\alpha\in\{p,v,T\}} \rho_K^\alpha \omega_K^\alpha \right\} + R_h, \tag{9.5}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

where the residual terms and weights are defined by

$$\begin{aligned}
\rho_K^p &:= \|R^p(u_h)\|_K,\\
\omega_K^p &:= \|r-r_h\|_K + \delta_K^p \|S^p(u_h)(r-r_h)\|_K,\\
\rho_K^v &:= \|R^v(u_h)\|_K + h_K^{-1/2}\|r^v(u_h)\|_{\partial K},\\
\omega_K^v &:= \|w-w_h\|_K + h_K^{1/2}\|w-w_h\|_{\partial K} + \delta_K^v \|S^v(u_h)(w-w_h)\|_K,\\
\rho_K^T &:= \|R^T(u_h)\|_K + h_K^{-1/2}\|r^T(u_h)\|_{\partial K},\\
\omega_K^T &:= \|S-S_h\|_K + h_K^{1/2}\|S-S_h\|_{\partial K} + \delta_K^T \|S^T(u_h)(S-S_h^T)\|_K,
\end{aligned}$$

with a suitable approximation $\{r_h, w_h, S_h\} \in V_h$ to z. The remainder term R_h can be bounded as in Proposition 8.1 for incompressible flow.

The details of the proof are omitted. The argument is similar to that used in the proof of Proposition 8.1 for incompressible flow. This analysis assumes for simplicity that viscosity μ and heat conductivity κ are determined by the reference temperature T_0, *i.e.*, these quantities are not included in the linearization process. This is justified because of their relatively small variation with T. Furthermore, the explicit dependence of the stabilization parameters δ_K on the discrete solution u_h is neglected.

The weights in the *a posteriori* error bound (9.5) are again evaluated by solving the adjoint problem numerically on the current mesh and approximating the exact adjoint solution z by patch-wise higher-order interpolation of its computed approximation $\tilde{z}_h$, as described above. This technique shows sufficient robustness and does not require the determination of any interpolation constant. For illustration, we state the strong form of the adjoint problem (suppressing terms related to the least-squares stabilization) used in our test computations:

$$\begin{aligned}
\nabla\cdot w &= j^p,\\
-\rho(v\cdot\nabla)w - \nabla\cdot\mu\nabla w - \rho\nabla r &= j^v,\\
T^{-1}v\cdot\nabla r + T^{-2}v\cdot\nabla T\cdot r - \rho v\cdot\nabla S - \ c_p^{-1}\nabla\cdot(\lambda\nabla S) - (Df^T)_T S &= j^T,
\end{aligned}$$

where $\{j^p, j^v, j^T\}$ is a suitable function representation of the error functional $J(\cdot)$. This system is closed by appropriate boundary conditions.

Numerical results (*from Becker, Braack and Rannacher* (2000*a*))
The first example (without chemistry) is a 2D benchmark 'heat-driven cavity' (for details see Figure 9.1 and Quere and Paillere (2000) or Becker *et al.* (2000*a*)). Here, the flow is confined to a square box with side length $L=1$ and is driven by a temperature difference $T_h - T_c = 2\varepsilon T_0$ between the left ('hot') and the right ('cold') wall under the action of gravity g in the y-direction.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

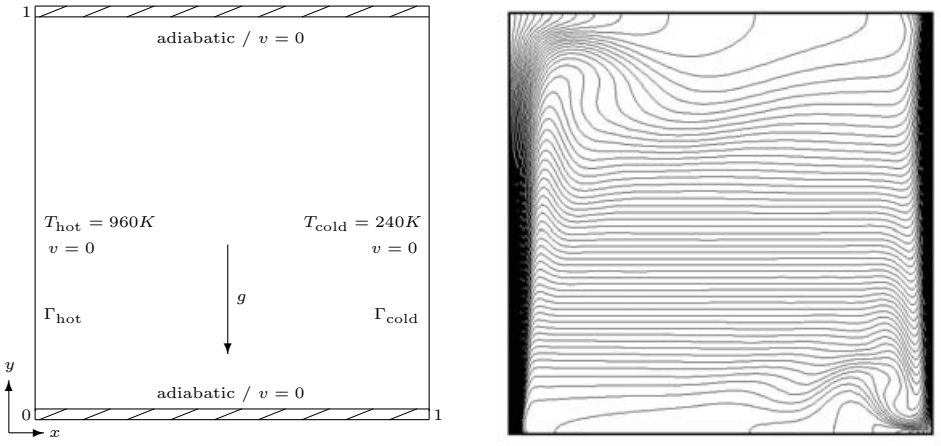


Fig. 9.1. Configuration of the heat-driven cavity problem and plot of computed temperature isolines

For the viscosity μ, we use Sutherland's law,

$$\mu(T) \;=\; \mu^*\left(\frac{T}{T^*}\right)^{1/3}\frac{T^*+S}{T+S},$$

with the Prandtl number $\mathrm{Pr} = 0.71$, $T^* = 273\,K$, $\mu^* = 1.68{\cdot}10^{-5}\mathrm{kg/ms}$, and $S := 110.5\,K$. Further, the heat conductivity is $\kappa(T) = \mu(T)/\mathrm{Pr}$. In the stationary case the thermodynamic pressure is defined by

$$P_{\text{th}} \;=\; P_0\left(\int_\Omega T_0^{-1}\,\mathrm{d}x\right)\left(\int_\Omega T^{-1}\,\mathrm{d}x\right)^{-1},$$

where $T_0 = 600\,K$ is a reference temperature and $P_0 = 101,325\,\mathrm{Pa}$. Accordingly, the Rayleigh number is determined by

$$\mathrm{Ra} \;=\; \mathrm{Pr}\,g\left(\frac{\rho_0 L}{\mu_0}\right)^2\frac{T_h-T_c}{T_0} \;\approx\; 10^6, \qquad \varepsilon \;=\; \frac{T_h-T_c}{T_h+T_c} \;=\; 0.6,$$

where $\mu_0 := \mu(T_0)$, $\rho_0 := P_0/RT_0$, and $R = 287\,\mathrm{J/kgK}$.

In this benchmark, one of the quantities to be computed is the average Nusselt number along the cold wall defined by

$$J(T) := c\int_{\Gamma_{\text{cold}}} \kappa\partial_n T\,\mathrm{d}s, \quad c := \frac{\mathrm{Pr}}{2\mu_0 T_0 \varepsilon}.$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Table 9.1. Results of computing the Nusselt number by using the 'energy error' indicator (left) and the weighted error indicator (right)

N	$\langle Nu\rangle_c$	$J(e)$	N	$\langle Nu\rangle_c$	$J(e)$	η_ω	I_{eff}
524	-9.09552	4.1e-1	523	-8.86487	1.8e-1	3.4e-1	1.9
945	-8.67201	1.5e-2	945	-8.71941	3.3e-2	1.5e-1	4.5
1708	-8.49286	1.9e-1	1717	-8.66898	1.8e-2	6.6e-2	3.7
3108	-8.58359	1.0e-1	5530	-8.67477	1.2e-2	2.1e-2	1.8
5656	-8.59982	8.7e-2	9728	-8.68364	3.0e-3	1.1e-2	3.7
18204	-8.64775	3.9e-2	17319	-8.68744	8.5e-4	6.3e-3	7.4
32676	-8.66867	1.8e-2	31466	-8.68653	6.9e-5	3.7e-3	53.6
58678	-8.67791	8.7e-3	56077	-8.68653	6.7e-5	2.1e-3	31.3
79292	-8.67922	7.4e-3	78854	-8.68675	1.5e-4	1.5e-3	10.0

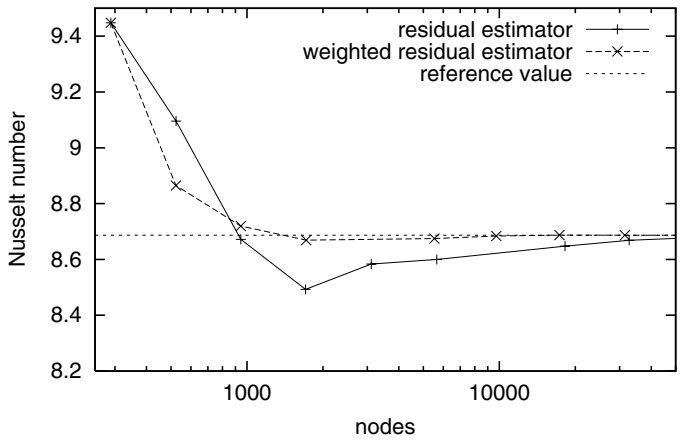


Fig. 9.2. Results of computing the Nusselt number by using a heuristic 'residual indicator' (symbol '+') and the 'weighted residual estimator' (symbol '×')

The results shown in Table 9.1 and Figure 9.2 demonstrate the mesh efficiency of the 'weighted' error indicator compared to a heuristic 'energy error' indicator similar to the one discussed in Section 8 for the incompressible flow case. The superiority of the 'weighted' indicator is particularly evident if higher solution accuracy is required; see Table 9.1. As usual in problems of such complex structure, the *a posteriori* bound $\eta(u_h)$ tends to

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

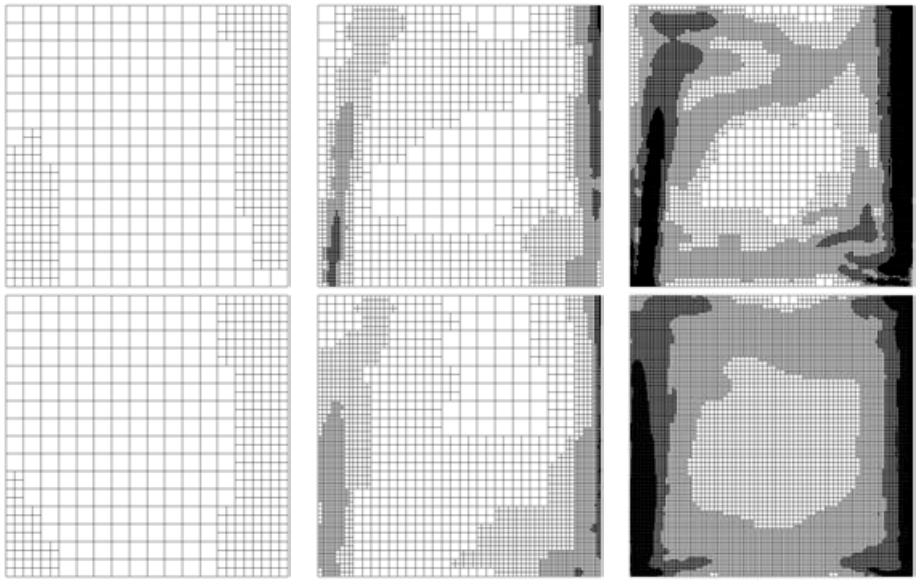


Fig. 9.3. Sequences of refined meshes obtained by the 'energy error' indicator (upper row, $N = 524, 5656, 58678$) and the weighted error estimator (lower row, $N = 523, 5530, 56077$)

over-estimate the true error but it appears to be 'reliable'. A collection of refined meshes produced by the two error indicators is shown in Figure 9.3.

Remark 9.1. The most important feature of the *a posteriori* error estimate (9.11) is that the local cell residuals related to the various physical effects governing flow and transfer of temperature are systematically weighted according to their impact on the error quantity to be controlled. For illustration, let us consider control of the mean velocity

$$J(v) = |\Omega|^{-1} \int_\Omega v \, \mathrm{d}x.$$

Then, in the adjoint problem the right-hand sides j^p and j^T vanish, but because of the coupling of the variables all components of the adjoint solution $z = \{r, w, S\}$ will be nonzero. Consequently, the error term to be controlled is also affected by the cell residuals of the mass-balance and the energy equation. This sensitivity is quantitatively represented by the weights involving r and S.

9.2. Chemically reactive flow in a 'flow reactor'

Next, we extend the 'low-Mach-number' flow model (9.1) by including chemical reactions. In this case the equations of mass, momentum and energy

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

conservation are supplemented by the equations of species mass conservation:

$$\nabla\cdot v - T^{-1}v\cdot\nabla T - M^{-1}v\cdot\nabla M = 0, \tag{9.6}$$

$$(\rho v\cdot\nabla)v + \nabla\cdot\tau + \nabla p = \rho f_e, \tag{9.7}$$

$$\rho v\cdot\nabla T - c_p^{-1}\nabla\cdot(\lambda\nabla T) = c_p^{-1}f_t(T,w), \tag{9.8}$$

$$\rho v\cdot\nabla w_i - \nabla\cdot(\rho D_i\nabla w_i) = f_i(T,w), \quad i=1,\dots,n. \tag{9.9}$$

The gas law takes the form

$$\rho = \frac{P_* M}{RT}, \tag{9.10}$$

with the mean molar mass $M := (\sum_{i=1}^n w_i/M_i)^{-1}$ and the species mole masses M_i (see, *e.g.*, Braack and Rannacher (1999)).

Owing to exponential dependence on temperature (Arrhenius' law) and polynomial dependence on w, the source terms $f_i(T,w)$ are highly nonlinear. In general, these zero-order terms lead to a coupling between all chemical species mass fractions. For robustness the resulting system of equations is to be solved by an implicit and fully coupled process that uses strongly adapted meshes. The discretization of the full flow system again uses continuous Q_1-finite elements for all unknowns and employs least-squares stabilization for the velocity-pressure coupling as well as for all the transport terms. We do not state the corresponding discrete equations since they have the same structure as seen before. The derivation of the related (linearized) adjoint problem corresponding to some (linear) functional $J(\cdot)$ and the resulting *a posteriori* error estimates follows the same line of argument. Again, we use a simplified Jacobian of the coupled system in setting up the adjoint problem (for details see Braack and Rannacher (1999). The resulting *a posteriori* error estimator has the same structure as that in Proposition 9.1, only that additional terms occur due to the balance equations for the chemical species:

$$|J(e)| \;\approx\; \eta_\omega(u_h) := \sum_{K\in\mathbb{T}_h}\left\{\sum_{\alpha\in\{p,v,T,w_i\}}\rho_K^\alpha\omega_K^\alpha\right\}, \tag{9.11}$$

with the chemistry-related additional residual terms and weights

$$\rho_K^{w_i} := \|R^{w_i}(u_h)\|_K + h_K^{-1/2}\|r^{w_i}(u_h)\|_{\partial K}, \quad i=1,\dots,n,$$

$$\omega_K^{w_i} := \|v_i - v_{i,h}\|_K + h_K^{1/2}\|v_i - v_{i,h}\|_{\partial K} + \delta_K^{w_i}\|S^{w_i}(u_h)(v_i - v_{i,h})\|_K.$$

The weights in the *a posteriori* error bound (9.11) are again computed in the same way as described above.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Numerical results (*from Waguet* (2000))
As a practically relevant example, we consider a chemical flow reactor (see Figure 9.4) for determining the reaction velocity of the heterogeneous relaxation of vibrationally excited hydrogen and its energy transfer in collisions with deuterium ('slow' chemistry),

$$H_2^{(1)} \rightarrow_{\text{wall}} H_2^{(0)}, \qquad H_2^{(1)} + D_2^{(0)} \rightarrow H_2^{(0)} + D_2^{(1)}.$$

The quantity to be computed is the CARS signal (coherent anti-Stokes Raman spectroscopy)

$$J(c) = \kappa \int_{-R}^{R} \sigma(s) c(r-s)^2 \, \mathrm{d}s,$$

where $c(r)$ is, for example, the concentration of $D_2^{(1)}$ along the line of the laser measurement (see Segatz *et al.* (1996)).

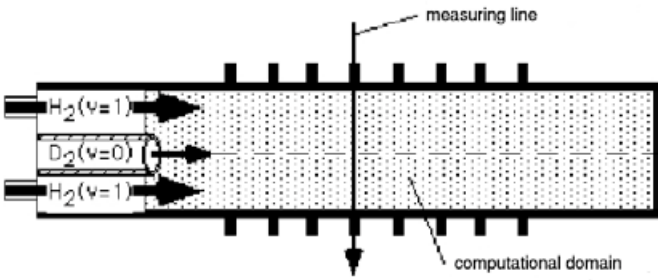


Fig. 9.4. Configuration of flow reactor

Table 9.2 and Figure 9.5 (overleaf) show results obtained by the DWR method for computing the mass fraction of $D_2^{(1)}$ and $D_2^{(0)}$. The comparison is against computations on heuristically refined tensor-product meshes. We observe higher accuracy on the systematically adapted meshes: in particular, monotone convergence of the quantities is achieved.

Notes and references
More complex combustion processes (gas-phase reactions: 'fast' chemistry) like the flame-sheet model of the Bunsen burner (laminar methane flame), the 3-species ozone recombination, or a model with detailed chemistry of (laminar) combustion in a methane burner involving 17 species and 88 reactions have been treated using the DWR method by Braack (1998); see also Becker, Braack, Rannacher and Waguet (1999), Braack and Rannacher (1999), and Becker *et al.* (2000*a*).

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Table 9.2. Some results of simulation for the relaxation experiment on hand-adapted (left) and on automatic-adapted (right) meshes

Heuristic refinement				Adaptive refinement			
L	N	$D_2^{(\nu=0)}$	$D_2^{(\nu=1)}$	L	N	$D_2^{(\nu=0)}$	$D_2^{(\nu=1)}$
2	481	0.742228	0.002541	2	244	0.738019	0.004020
3	1793	0.780133	0.002531	3	446	0.745037	0.002600
4	1923	0.782913	0.002729	4	860	0.756651	0.002010
5	2378	0.785116	0.001713	5	1723	0.780573	0.001390
6	3380	0.791734	0.001162	6	3427	0.785881	0.001130
7	5374	0.791627	0.001436	7	7053	0.799748	0.001090

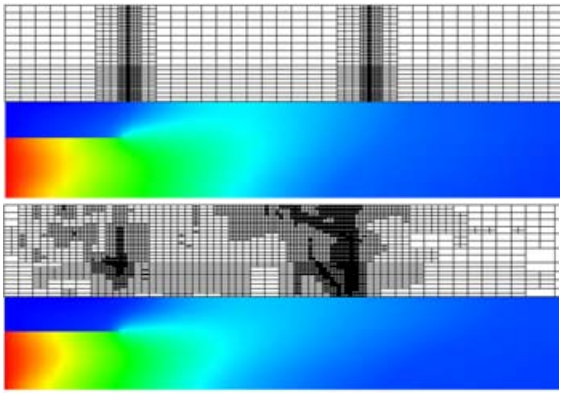


Fig. 9.5. Mass fraction of $D_2^{(1)}$ in the flow reactor computed on a tensor-product mesh (top) and on a locally adapted mesh (bottom)

10. Application to elasto-plasticity

We present an example of the application of the DWR method in solid mechanics, particularly in elasto-plasticity theory. The problem to be solved is strongly nonlinear with a nonlinearity not everywhere differentiable. Nevertheless, the straightforward application of the DWR method to this situation gives satisfactory results. This indicates that the method is more robust in practice than one might expect from theory.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

10.1. The Hencky model of perfect plasticity

The fundamental problem in the static deformation theory of linear-elastic perfect plastic material (the so-called *Hencky* model) reads

$$\begin{gathered}\nabla\cdot\sigma = -f, \quad \varepsilon(u) \;=\; A:\sigma + \lambda \quad \text{in } \Omega,\\ \lambda:(\tau-\sigma) \le 0, \quad \text{for all } \; \tau \in \Sigma := \{\tau, \mathcal{F}(\tau) \le 0\},\\ u = 0 \;\text{ on } \Gamma_D, \quad \sigma\cdot n = g \;\text{ on } \Gamma_N,\end{gathered} \tag{10.1}$$

where σ and u are the stress tensor and displacement vector, respectively, while λ denotes the plastic growth. This system describes the deformation of an elasto-plastic body occupying a bounded domain $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3) which is fixed along a part Γ_D of its boundary $\partial\Omega$, under the action of a body force with density f and a surface traction g along $\Gamma_N = \partial\Omega\backslash\Gamma_D$. The displacement u is assumed to be small in order to neglect geometric nonlinear effects, so that the strain tensor can be written in the form $\varepsilon(u) = \frac{1}{2}(\nabla u + \nabla u^T)$. We assume a linear-elastic isotropic material law, that is, the material tensor A is given in the form $A = C^{-1}$, where

$$\sigma = C\varepsilon := 2\mu\varepsilon^D + \kappa\,\mathrm{tr}(\varepsilon)I,$$

with material-dependent constants $\mu > 0$ and $\kappa > 0$. The plastic behaviour follows the von Mises flow rule

$$\mathcal{F}(\sigma) := |\sigma^D| - \sigma_0 \le 0,$$

with some $\sigma_0 > 0$. Here, $\varepsilon^D := \varepsilon - \frac{1}{d}\mathrm{tr}(\varepsilon)I$ and $\sigma^D := \sigma - \frac{1}{d}\mathrm{tr}(\sigma)I$ are the deviatoric parts of ε and σ, respectively. For setting up the *primal* variational formulation of problem (10.1), we introduce the Hilbert space $V := \{u \in H^1(\Omega)^d,\ u_{|\Gamma_D} = 0\}$. Further, we define the nonlinear material function

$$C(\varepsilon) := \begin{cases} C\varepsilon, & \text{if } \; |2\mu\varepsilon^D| \le \sigma_0,\\ \dfrac{\sigma_0}{|\varepsilon^D|}\varepsilon^D + \kappa\,\mathrm{tr}(\varepsilon)I, & \text{if } \; |2\mu\varepsilon^D| > \sigma_0. \end{cases}$$

The tensor-function $C(\cdot)$ is globally only Lipschitz and not differentiable along the yield surface $\{\tau \in \mathbb{R}^{d\times d}_{\mathrm{sym}},\ |2\mu\tau^D| = \sigma_0\}$. Below, we will use the piecewise differential $C'(\tau)$ defined by

$$C'(\tau)\varepsilon := \begin{cases} C\varepsilon, & \text{if } \; |2\mu\tau^D| \le \sigma_0,\\ \dfrac{\sigma_0}{|\tau^D|}\Big\{I - \dfrac{(\tau^D)^T\tau^D}{|\tau^D|^2}\Big\}\varepsilon^D + \kappa\,\mathrm{tr}(\varepsilon)I, & \text{if } \; |2\mu\tau^D| > \sigma_0. \end{cases}$$

Then, we seek a displacement $u \in V$, satisfying

$$A(u;\varphi) = F(\varphi), \quad \text{for all } \varphi \in V, \tag{10.2}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

with the semilinear and linear forms

$$A(u;\varphi) := (C(\varepsilon(u)),\varepsilon(\varphi)), \qquad F(\varphi) := (f,\varphi) + (g,\varphi)_{\Gamma_N}.$$

The finite element approximation of problem (10.2) seeks $u_h \in V_h$ such that

$$A(u_h;\varphi_h) = F(\varphi_h), \quad \text{for all } \varphi_h \in V_h, \tag{10.3}$$

where V_h is the finite element space of bilinear shape functions as described above. Having computed the displacement u_h, we obtain a corresponding stress by $\sigma_h := C(\varepsilon(u_h))$. Details of the solution process can be found in Rannacher and Suttmeier (1999). In practice, the exact evaluation of $C(\cdot)$ is difficult, so the formal scheme (10.2) requires some modification. This may be accomplished by storing information only at Gaussian points (2×2-Gauss points for Q_1-elements) at which $C(\cdot)$ is evaluated exactly. However, in the following, it is always assumed that the discretization is realized in its ideal form without numerical integration. The nonlinear algebraic problem (10.3) is approximated by a modified Newton iteration as described in Section 6. The linear subproblems are solved by the CR method with multigrid acceleration (see Suttmeier (1996) for more details of the algebraic solution techniques).

Now, we turn to the *a posteriori* error analysis. Given a (linear) functional $J(\cdot)$ on V, we consider the corresponding linearized adjoint problem

$$A'(u;\varphi,z) = J(\varphi), \quad \text{for all } \varphi \in V, \tag{10.4}$$

where in the tangent bilinear form

$$A'(u;\varphi,z) = (C'(\varepsilon(u))\varepsilon(\varphi),\varepsilon(z)),$$

the matrix operator $C'(\varepsilon(u_h)$ is defined as above. As before, we define the equation residual

$$R_{|K} := f + \nabla\cdot\sigma_h, \quad \sigma_h := C(\varepsilon(u_h)),$$

and the edge residuals

$$r_{|\Gamma} := \begin{cases} \frac{1}{2}n\cdot[\sigma_h], & \text{if } \Gamma \subset \partial K\backslash\partial\Omega, \\ 0, & \text{if } \Gamma \subset \Gamma_D, \\ n\cdot\sigma_h - g, & \text{if } \Gamma \subset \Gamma_N. \end{cases}$$

With this notation, we obtain from Proposition 2.3 the following result.

Proposition 10.1. (Rannacher and Suttmeier 1998) For the approximation of the Hencky model (10.1) by the finite element scheme (10.3), we have the following *a posteriori* estimate for the error $e := u - u_h$:

$$|J(e)| \;\approx\; \eta_\omega(u_h) := \sum_{K\in\mathbb{T}} \rho_K\,\omega_K + R, \tag{10.5}$$

modulo a remainder term R that is formally quadratic in e, where the local

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

residual terms and weights are defined by

$$\rho_K := \|R(u_h)\|_K + h_K^{-1/2}\|r(u_h)\|_{\partial K}, \qquad \omega_K := \|z-\varphi_h\|_K + h_K^{1/2}\|z-\varphi_h\|_{\partial K},$$

with a suitable approximation $\varphi_h \in V_h$ to the adjoint solution z.

Proof. From Proposition 2.3, we have the abstract error representation

$$J(e) \;=\; A(x_h; z-y_h) + R,$$

with a remainder term R, which is not further considered here. Using the definition of the semilinear form $A(\cdot;\cdot)$, this can be written in the concrete form

$$J(e) \;\approx\; \sum_{K\in\mathbb{T}_h} \big\{(R(u_h), z-z_h)_K - (r(u_h), z-z_h)_{\partial K}\big\},$$

from which we immediately conclude the asserted estimate. □

The *a posteriori* error estimator (10.5) may be evaluated by either one of the strategies described above. For simplicity, the linearization of the adjoint problem uses a decomposition of the mesh domain $\Omega_h = \cup\{K \in \mathbb{T}_h\}$ into its 'elastic' component Ω_h^e and 'plastic' component Ω_h^p defined by

$$\Omega_h^e := \cup\{K\in\mathbb{T}_h,\, |2\mu\varepsilon^D{}_{|K}| \le \sigma_0\}, \qquad \Omega_h^p := \Omega \setminus \Omega_h^e.$$

Then, we define on each cells $T \in \mathbb{T}_h$:

$$C_h'(\tau)(\varepsilon)_{|K} := \begin{cases} C\varepsilon, & \text{if } K \subset \Omega_h^e, \\ \dfrac{\sigma_0}{|\tau^D|}\left\{I - \dfrac{(\tau^D)^K \tau^D}{|\tau^D|^2}\right\}\varepsilon^D + \kappa\,\mathrm{tr}(\varepsilon)I, & \text{if } K \subset \Omega_h^p. \end{cases}$$

Using this notation, the adjoint solution z is approximated by the solution $\tilde z_h \in V_h$ of the discretized adjoint problem

$$(\varepsilon(\varphi_h), C_h'(\varepsilon(u_h))^*\varepsilon(\tilde z_h)) = J(\varphi_h), \quad \text{for all } \varphi_h \in V_h, \tag{10.6}$$

The evaluation of the coefficient $C_h'(\varepsilon(u_h))^*$ on cells in the elastic–plastic transition zone is usually done by simple numerical integration. This may appear to be a rather crude approximation, but it works in practice. The reason may be that the critical situation only occurs in cells intersecting the elastic–plastic transition zone, which is a lower-dimensional surface. The weights ω_K may then again be approximated as described in Section 5. We emphasize that the computation of the adjoint solution requires us to solve only *linear* problems and normally only amounts to a small fraction of the total cost within a Newton iteration for the nonlinear problem. We will compare the weighted error estimator $\eta_\omega(u_h)$ with two heuristic indicators for the stress error $e_\sigma := \sigma - \sigma_h$.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

- The heuristic ZZ-error indicator of Zienkiewicz and Zhu (1987) (see, *e.g.*, Ainsworth and Oden (1997*b*)) uses the idea of higher-order stress recovery by local averaging,

$$\|e_\sigma\| \approx \eta_{ZZ}(u_h) := \left(\sum_{K\in\mathbb{T}_h} \|M_h\sigma_h - \sigma_h\|_K^2 \right)^{1/2}, \tag{10.7}$$

  where $M_h\sigma_h$ is a local (super-convergent) approximation of σ.

- The heuristic energy error estimator of Johnson and Hansbo (1992*a*, 1992*b*) and Hansbo (200x) is based on the decomposition $\Omega = \Omega_h^e \cup \Omega_h^p$:

$$\|e_\sigma\| \approx \eta_E(u_h) := c_i \left(\sum_{K\in\mathbb{T}_h} \eta_K^2 \right)^{1/2}, \tag{10.8}$$

  where

$$\eta_K^2 := \begin{cases} h_K^2 \max_K |R(u_h)|^2, & \text{if } K \in \Omega_h^e, \\ h_K^{-1} \max_K |C\varepsilon(u_h) - \mathcal{M}_h C\varepsilon(u_h)| \int_K |R(u_h)|\,\mathrm{d}x, & \text{if } K \in \Omega_h^p. \end{cases}$$

Numerical results (*from Rannacher and Suttmeier* (1998))
The approach described above is applied for a typical model problem in elasto-plasticity employing the two-dimensional 'plane strain' model (for more details and further examples see Rannacher and Suttmeier (1997)). A square disc with a crack is subjected to a constant boundary traction acting on the upper boundary (see Figure 10.1). Along the right-hand side and the lower boundary the disc is fixed and the remaining part of the boundary (including the crack) is left free. This problem is interesting as its solution develops a singularity at the tip of the crack where a strong stress concentration occurs which causes plastification locally. The material parameters chosen are those commonly used for aluminium: $\mu = 80193.80\,\mathrm{Nmm}^{-2}$, $\kappa = 164206\,\mathrm{Nmm}^{-2}$, and $\sigma_0 = \sqrt{2/3}\,450$.

We want to compute the mean normal stress over the clamped boundary,

$$J(\sigma) = \int_{\Gamma_u} n^T \sigma n \,\mathrm{d}s. \tag{10.9}$$

Since this functional is irregular, it is regularized as

$$J_\varepsilon(\tau) := \frac{1}{|\Gamma_\varepsilon|} \int_{\Gamma_\varepsilon} n^T \tau n \,\mathrm{d}x, \quad \Gamma_\varepsilon = \{x \in \mathbb{R}^2, \mathrm{dist}(x, \Gamma_D) < \tfrac{1}{2}\varepsilon\}, \tag{10.10}$$

with $\varepsilon = \mathrm{TOL}$. A reference solution σ_{ref} is computed on a fine mesh with about 200,000 cells for determining the relative error and the corresponding

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

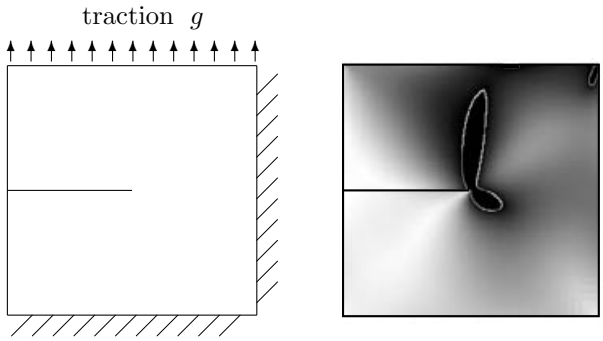


Fig. 10.1. Geometry of the test problem 'square disc with crack' and plot of $|\sigma^D|$ (plastic regions black) computed on a mesh with $N \approx 64,000$ cells

Table 10.1. Results obtained by the weighted *a posteriori* error estimator

N	$J(\sigma_h)$	E_{rel}	I_{eff}
1000	2.2224e+02	1.5757e-02	1.7
2000	2.2344e+02	1.0430e-02	2.1
4000	2.2405e+02	7.7387e-03	1.7
8000	2.2475e+02	4.6647e-03	1.6
16000	2.2532e+02	2.1360e-03	1.8
∞	2.2580e+02		

'effectivity index' (used here for expressing over-estimation) defined by

$$E_{\text{rel}} := \left| \frac{J_\varepsilon(\sigma_h - \sigma_{\text{ref}})}{J_\varepsilon(\sigma_{\text{ref}})} \right|, \quad I_{\text{eff}} := \left| \frac{\eta_\omega(u_h, \sigma_h)}{J_\varepsilon(\sigma_h - \sigma_{\text{ref}})} \right|.$$

In this case the weighted error estimator turns out to be rather sharp even on relatively coarse meshes; see Table 10.1. This indicates that the strategy of evaluating the weights ω_K computationally also works for the present nonlinear problem with nonsmooth nonlinearity. Further, as in the linear case, we obtain more economical meshes than by the other heuristic error estimators (Figure 10.2).

Remark 10.1. In the plastic region the material behaviour will be almost incompressible, which causes stability problems of the discretization based on the formulation (10.2). In order to cope with this problem, a stabilized

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

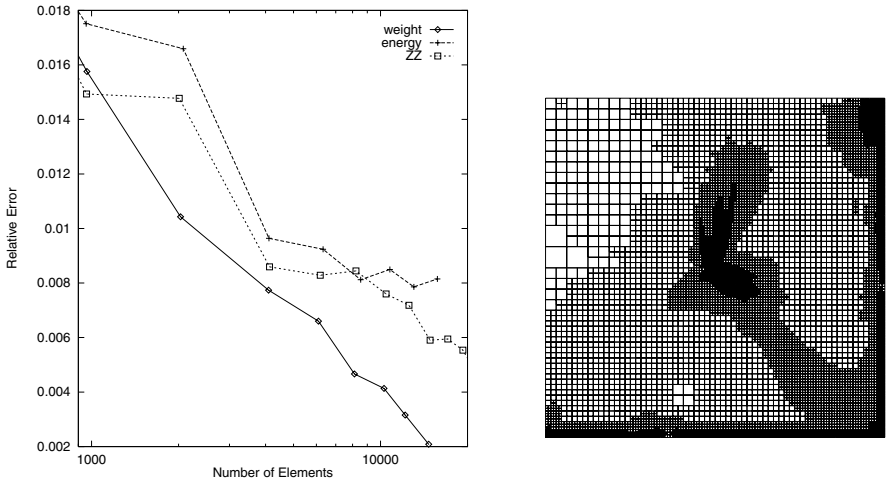


Fig. 10.2. Relative error for $J(\sigma)$ on grids based on the different estimators (left) and structure of optimized grid with $N \approx 8100$ (right)

finite element discretization may be employed by introducing an auxiliary 'pressure' variable (see Suttmeier (1998)).

Notes and references

The adaptive finite element method described above for solving the *stationary* Hencky problem in perfect plasticity has been extended in Rannacher and Suttmeier (1999, 200x) to the *quasi-stationary* Prandtl–Reuss model. Here, the time/load stepping is done by the backward Euler scheme and it is shown that the resulting incremental errors in each time/load step can accumulate at most linearly. The incorporation of the elasto-plasticity problem into the general framework of the DWR method relies on its reformulation as a nonlinear variational *equation*. For the application of duality techniques for *a posteriori* error estimation in the direct approximation of variational *inequalities*, we refer to Johnson (1992) and particularly to the recent papers of Blum and Suttmeier (1999, 2000) and Suttmeier (2001), where a variant of the DWR method is used.

11. Application to radiative transfer

In this section, we consider an example from astrophysics. The continuum model for the transfer of light including scattering is an integro-differential equation for the intensity as a function of time, frequency, space and direction. In astrophysical applications particular difficulties arise by the strong heterogeneity of the coefficients in the model. This type of problem poses

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

highest requirements on numerical simulation, and adaptive discretization is mandatory for achieving accurate results. This example is intended to demonstrate that the DWR method can directly be applied to finite element discretization of this nonstandard model.

11.1. The monochromatic radiative transfer equation

The emission of light of a certain wavelength from a cosmic source is described by the (monochromatic) 'radiative transfer equation'

$$\theta\cdot\nabla_x u + (\kappa+\sigma)u - \sigma\int_{S_2} P(\theta,\theta')u\,\mathrm{d}\theta' \;=\; B \quad \text{in } \Omega\times S_2, \tag{11.1}$$

for the radiation intensity $u = u(x,\theta)$. Here, $x\in\Omega\subset\mathbb{R}^3$ is a bounded domain and $\theta\in S_2$ the unit-sphere in $\mathbb{R}^3$. The usual boundary condition is $u=0$ at the 'inflow' boundary $\Gamma_{\mathrm{in},\theta} = \{x\in\partial\Omega, n\cdot\theta<0\}$. The absorption and scattering coefficients $\kappa(x)>0$, $\sigma(x)>0$, the redistribution kernel $P(\theta,\theta')\geq 0$, and the source term B (Planck function) are given. In applications these functions exhibit strong variations (several orders of magnitude) in space, which requires the use of locally refined meshes. Further, because of these coefficient irregularities, quantitative *a priori* information about the properties of the radiation operator and the solution's regularity are not available (see Kanschat (1996, 1998)).

The starting point of a numerical solution of the radiative transfer problem (11.1) is again its variational formulation. To this end, we introduce the solution space

$$V := \left\{v\in L^2(\Omega\times S_2),\ \theta\cdot\nabla_x u\in L^2(\Omega\times S_2)\right\}$$

and the operators

$$T_\theta u := \theta\cdot\nabla_x u, \qquad \Sigma u := \sigma u - \sigma\int_{S_2} P(\theta,\theta')u\,\mathrm{d}\theta'.$$

Then, we seek $u\in V$ satisfying

$$A(u,\varphi) = F(\varphi), \quad \text{for all } \varphi\in V, \tag{11.2}$$

with the linear forms

$$A(u,\varphi) := (T_\theta u+\kappa u+\Sigma u,\varphi)_{\Omega\times S_2} + (\theta\cdot nu,\varphi)_{\Gamma_{in,\theta}}, \qquad F(\varphi) := (B,\varphi)_{\Omega\times S_2}.$$

Notice that, in this formulation, the 'inflow' boundary conditions are incorporated weakly. The existence of unique solutions of (11.2) can be inferred by abstract functional analytic arguments. The discretization of this problem uses standard (continuous) Q_1-finite elements in $x\in\Omega$, on meshes $\mathbb{T}_h = \{K\}$ with local width h_K, and (discontinuous) P_0-finite elements in $\theta\in S_2$, on meshes $\mathbb{D}_k = \{\Delta\}$ of uniform width k_Δ. Again, streamline diffusion as described in Section 3.3 is employed in order to stabilize the transport

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

term $\theta \cdot \nabla_x u$. The discretization of the ordinate space S_2 is equivalent to using the midpoint quadrature formula on the integral operator Σ. The x-mesh is adaptively refined, while the θ-mesh is kept uniform (suggested by *a priori* error analysis).

Let $V_h \subset V$ be the finite element subspaces. The Galerkin finite element approximations $u_h \in V_h$ are defined by

$$A_h(u_h, \varphi_h) = F_h(\varphi_h), \quad \text{for all } \varphi_h \in V_h, \tag{11.3}$$

with the stabilized forms

$$A_h(u_h, \varphi_h) := A(u_h, \varphi_h + \delta T_\theta \varphi_h), \quad F(\varphi_h) := F(\varphi_h + \delta T_\theta \varphi_h),$$

and some mesh-dependent parameter $\delta \sim h \min\{\kappa^{-1}, \sigma^{-1}\}$. Clearly (11.3) is also satisfied by the exact solution u, such that Galerkin orthogonality again holds with respect to the form $A_h(\cdot, \cdot)$.

The refinement process is organized as described before on the basis of a weighted *a posteriori* error estimate for the target quantity $J(u)$. To this end, we consider the following adjoint problem for $z \in V$:

$$A_h(\varphi, z) = J(\varphi), \quad \text{for all } \varphi \in V, \tag{11.4}$$

which represents a modified radiative transport problem with source term $J(\cdot)$. The solvability of (11.4) follows from the general theory as for the original problem (11.2). From Proposition 2.6, we then obtain the following result.

Proposition 11.1. (Kanschat 1996) Let z be the solution of the adjoint problem (11.4). Then, for the finite element approximation (11.3) of the radiative transfer problem (11.1), we have the following *a posteriori* estimate:

$$|J(e)| \le \eta_\omega(u_h) := \sum_{\Delta \in \mathbb{D}_k} \sum_{K \in \mathbb{T}_h} \{\rho_K^1 \omega_K^1 + \rho_K^2 \omega_K^2\}, \tag{11.5}$$

where the residual terms and weights are defined by

$$\begin{aligned}
\rho_K^1 &:= \|B - (T_\theta + \kappa I - \Sigma) u_h\|_{K \times \Delta}, \\
\omega_K^1 &:= \|z - \varphi_h\|_{K \times \Delta} + \delta_K \|T_\theta (z - \varphi_h)\|_{K \times \Delta}, \\
\rho_K^2 &:= h_K^{-1/2} \|\theta \cdot n u_h\|_{\partial K \cap \Gamma_{in,\theta}}, \\
\omega_K^2 &:= h_K^{1/2} \|z - \varphi_h\|_{\partial K \cap \Gamma_{in,\theta}},
\end{aligned}$$

with a suitable approximation $\varphi_h \in V_h$ to z.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

The 'weighted' estimator estimator η_ω will be compared to a heuristic L^2-error indicator of the form

$$\|e\|_{\Omega\times S_2} \le c_s c_i \left(\sum_{\Delta\in\mathbb{D}_k} \sum_{K\in\mathbb{T}_h} (h_K^2 + k_\Delta^2)\, \rho_K^2 \right)^{1/2}, \tag{11.6}$$

where the stability constant c_s is either computed by solving numerically the adjoint problem corresponding to the source term $\|e\|^{-1}_{\Omega\times S_2} e$, or simply set to $c_s = 1$. The 'interpolation constant' c_i is usually of moderate size $c_i \sim 0.1 - 1$. The inclusion of the scaling factor $h_K^2 + k_\Delta^2$ is only based on heuristic grounds since useful analytical *a priori* estimates for the adjoint problem (11.4) are not available.

Numerical results (*from Kanschat* (1996))
We consider a prototypical example from astrophysics. A satellite-based observer measures the light (at a fixed wavelength) emitted from a cosmic source hidden in a dust cloud. A sketch of this situation is shown in Figure 11.1. The measurement is compared with results of a (two-dimensional) simulation which assumes certain properties of the coefficients in the underlying radiative transfer model (11.1). Because of the distance to the source, only the mean value of the intensity emitted in the direction θ_{obs} can be measured. Thus, the quantity to be computed is

$$J(u) = \int_{\{n\cdot\theta_{\text{obs}}\ge 0\}} u(x, \theta_{\text{obs}})\, \mathrm{d}s,$$

where $\{n \cdot \theta_{\text{obs}} \ge 0\}$ is the outflow boundary of the computational domain $\Omega \times S_1$ ($\Omega \subset \mathbb{R}^2$ a square) containing the radiating object.

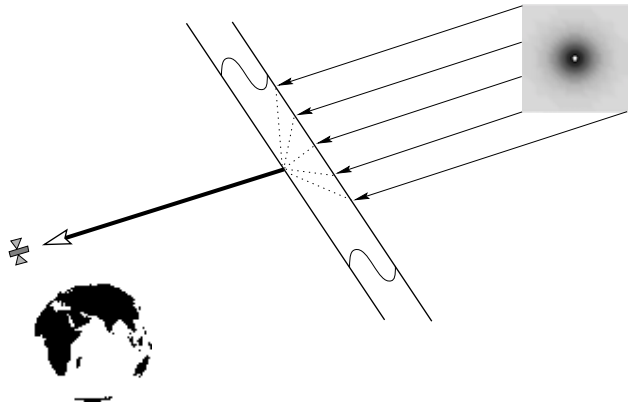


Fig. 11.1. Observer configuration of radiation emission

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Table 11.1. Results obtained by the (heuristic) L^2-error indicator and the weighted error estimator, the total number of unknowns being $N_{\text{tot}} = N_x \cdot 32$

	L^2-indicator		Weighted estimator			
L	N_x	$J(u_h)$	N_x	$J(u_h)$	$\eta_\omega(u_h)$	$\eta_\omega(u_h)/J(e)$
2	1105	0.210	1146	0.429	1.0804	8.62
3	2169	0.311	2264	0.461	0.7398	7.11
4	4329	0.405	4506	0.508	0.2861	3.94
5	8582	0.460	9018	0.555	0.1375	3.33
6	17202	0.488	18857	0.584	0.0526	2.39
7	34562	0.537	39571	0.599	0.0211	1.76
8	68066	0.551	82494	0.608	0.0084	1.41
∞				0.618		

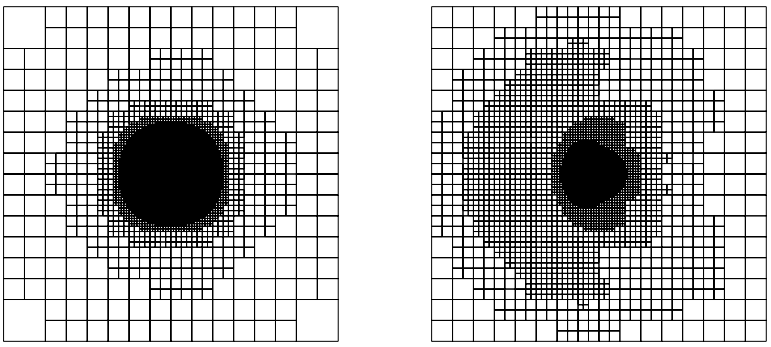


Fig. 11.2. Optimized meshes generated by the (heuristic) L^2-error indicator (left) and the weighted error estimator (right). The observer direction is to the left

The results in Table 11.1 demonstrate the superiority of the weighted error estimator over the heuristic global L^2-error indicator. The effect of the presence of the weights on the mesh refinement is shown in Figure 11.2.

Notes and references

Spatially three-dimensional radiative transfer problems in astrophysics can be solved with satisfactory accuracy *only* by systematically exploiting mesh adaptation and parallel computers. The application of the DWR method in

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

3D and the parallel implementation of the fully adaptive solution method has been accomplished by Kanschat (1996, 2000). A rigorous *a posteriori* error analysis can be found in Führer and Kanschat (1997), while the particular aspects of streamline diffusion in the context of radiative transport have been discussed by Kanschat (1998). For realistic astrophysical applications using the DWR see Wehrse, Meinkoehn and Kanschat (1999).

12. Application to optimal control

As the last example, we demonstrate the application of the DWR method in optimal control. This is a situation which is naturally suited to 'goal-oriented' *a posteriori* error estimation.

12.1. A linear model problem

We have chosen a very simple model problem with linear state equation in order to make the particular features of our approach clear. The state equation is given in terms of the mixed boundary value problem of the Helmholtz operator

$$\begin{aligned} -\Delta u + u &= f \quad \text{on } \Omega, \\ \partial_n u &= q \ \text{ on } \Gamma_C, \quad \partial_n u = 0 \ \text{ on } \partial\Omega \setminus \Gamma_C, \end{aligned} \tag{12.1}$$

defined on a bounded domain $\Omega \subset \mathbb{R}^2$ with boundary $\partial\Omega$. The control q acts on the boundary component Γ_C, while the observations $u_{|\Gamma_O}$ are taken on a component Γ_O; see Figure 12.1, on page 90. The cost functional is defined by

$$J(u,q) = \tfrac{1}{2}\|u - u_O\|^2_{\Gamma_O} + \tfrac{1}{2}\alpha\|q\|^2_{\Gamma_C}, \tag{12.2}$$

with a prescribed function u_O and a parameter $\alpha > 0$. We want to apply the general formalism of Section 2 to the Galerkin finite element approximation of this problem. This may be considered within the context of 'model reduction' in optimal control theory. First, we have to prepare the corresponding functional analytic setting. The functional of interest is the Lagrangian functional of the optimal control problem,

$$L(x) = J(u,q) + (\nabla u, \nabla z) + (u - f, z) - (q, z)_{\Gamma_C},$$

defined for triples $x = \{u, z, q\}$ in the Hilbert space $X := V \times V \times Q$, where $V := H^1(\Omega)$ and $Q := L^2(\Gamma_C)$. The Euler–Lagrange equations for stationary points $x = \{u, z, q\} \in X$ of $L(\cdot)$ are

$$L'(x; y) = 0, \quad \text{for all } y \in X, \tag{12.3}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

or, written in explicit form,

$$\begin{aligned}(\varphi^u, u-u_O)_{\Gamma_O} + (\nabla\varphi^u, \nabla z) + (\varphi^u, z) &= 0, \quad \text{for all } \varphi^u \in V,\\ (\nabla u, \nabla\varphi^z) + (u-f, \varphi^z) - (q, \varphi^z)_{\Gamma_C} &= 0, \quad \text{for all } \varphi^z \in V,\\ (z-\alpha q, \varphi^q)_{\Gamma_C} &= 0, \quad \text{for all } \varphi^q \in Q.\end{aligned} \tag{12.4}$$

The corresponding discrete approximations $x_h = \{u_h, z_h\, q_h\}$ are determined in the finite element space $X_h = V_h \times V_h \times Q_h \subset V$ by

$$\begin{aligned}(\psi_h, u_h-u_O)_{\Gamma_O} + (\nabla\psi_h, \nabla z_h) + (\psi_h, z_h) &= 0, \quad \text{for all } \psi_h \in V_h,\\ (\nabla u_h, \nabla\pi_h) + (u_h-f, \pi_h) - (q_h, \pi_h)_{\Gamma_C} &= 0, \quad \text{for all } \pi_h \in V_h,\\ (z_h-\alpha q_h, \chi_h)_{\Gamma_C} &= 0, \quad \text{for all } \chi_h \in Q_h.\end{aligned} \tag{12.5}$$

Here, the trial spaces V_h for the state and co-state variables are as defined above in Section 3 (isoparametric bilinear shape functions), and the spaces Q_h for the controls consist of traces on Γ_C of V_h-functions, for simplicity. Following the general formalism, we seek to estimate the error $e = \{e^u, e^z, e^q\}$ with respect to the Lagrangian functional $L(\cdot)$. Proposition 2.1 yields the following abstract *a posteriori* estimate for the error in the Lagrangian:

$$|L(x) - L(x_h)| \;\le\; \eta(x_h) := \inf_{y_h \in X_h} \tfrac{1}{2}|L'(x_h; x-y_h)|. \tag{12.6}$$

In the present case, the cost functional is quadratic and the state equation linear, so that the remainder term in (2.6) vanishes. Since $\{u, z, q\}$ and $\{u_h, z_h, q_h\}$ satisfy (12.4) and (12.5), respectively,

$$\begin{aligned}L(x) - L(x_h) \;&=\; J(u,q) + (\nabla u, \nabla z) + (u-f, z) - (q, z)_{\Gamma_C}\\ &\quad - J(u_h, q_h) - (\nabla u_h, \nabla z_h) - (u_h - f, z_h) + (q_h, z_h)_{\Gamma_C}\\ &=\; J(u,q) - J(u_h, q_h).\end{aligned}$$

Hence, error control with respect to the Lagrangian functional $L(\cdot)$ and the cost functional $J(\cdot)$ are equivalent. Now, evaluation of the abstract error bound (12.6) again employs splitting the integrals into the contributions by the single cells, cell-wise integration by parts and Hölder's inequality (for the detailed argument see Kapp (2000) and Becker, Kapp and Rannacher (2000*c*)). For the following analysis, we introduce the cell residuals

$$\begin{aligned}R^u(x_h)_{|K} &:=\; f + \Delta u_h - u_h,\\ R^z(x_h)_{|K} &:=\; \Delta z_h - z_h,\end{aligned}$$

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

and edge residuals

$$r^u(x_h)_{|\Gamma} := \begin{cases} \frac{1}{2} n \cdot [\nabla u_h], & \text{if } \Gamma \not\subset \partial\Omega, \\ \partial_n u_h, & \text{if } \Gamma \subset \partial\Omega \backslash \Gamma_C, \\ \partial_n u_h - q_h, & \text{if } \Gamma \subset \Gamma_C, \end{cases}$$

$$r^z(x_h)_{|\Gamma} := \begin{cases} \frac{1}{2} n \cdot [\nabla z_h], & \text{if } \Gamma \not\subset \partial\Omega, \\ \partial_n z_h, & \text{if } \Gamma \subset \partial\Omega \backslash \Gamma_O, \\ \partial_n z_h + u_h - u_O, & \text{if } \Gamma \subset \Gamma_O, \end{cases}$$

$$r^q(x_h)_{|\Gamma} := \begin{cases} z_h - \alpha q_h, & \text{if } \Gamma \subset \Gamma_C, \\ 0, & \text{if } \Gamma \not\subset \Gamma_C, \end{cases}$$

associated with the solution $x_h = \{u_h, z_h, q_h\}$ of the system (12.5), where the notation $[\cdot]$ has its usual meaning. With this notation, we obtain from the abstract *a posteriori* error estimate (12.6) the following result for the present situation.

Proposition 12.1. For the finite element discretization of the system (12.5), we have the *a posteriori* error estimate

$$\begin{aligned} |J(u,q) - J(u_h,q_h)| &\leq \eta_\omega(x_h) \\ &:= \sum_{K \in \mathbb{T}_h} \{\rho_K^u \, \omega_K^z + \rho_K^z \, \omega_K^u + \rho_K^q \, \omega_K^q\}, \end{aligned} \tag{12.7}$$

where the cell-wise residuals and weights are defined by

$$\begin{aligned} \rho_K^z &:= \|R^z(x_h)\|_K + h_K^{-1/2} \|r^z(x_h)\|_{\partial K}, \\ \omega_K^u &:= \|u - i_h u\|_K + h_K^{1/2} \|u - i_h u\|_{\partial K}, \\ \rho_K^u &:= \|R^u(x_h)\|_K + h_K^{-1/2} \|r^u(x_h)\|_{\partial K}, \\ \omega_K^z &:= \|z - i_h z\|_K + h_K^{1/2} \|z - i_h z\|_{\partial K}, \\ \rho_K^q &:= h_K^{-1/2} \|r^q(x_h)\|_{\partial K}, \\ \omega_K^q &:= h_K^{1/2} \|q - i_h q\|_{\partial K}, \end{aligned}$$

with suitable approximations $\{i_h u, i_h z, i_h q\} \in V_h \times V_h \times Q_h$.

Notice that the *a posteriori* error estimate (12.7) has very particular features. Its evaluation does not require the additional solution of an 'adjoint problem': the weights are rather generically obtained from the solution itself. The residuals of the state equation are weighted by the adjoint variable and, in turn, those of the adjoint equation by the primal variable. In this way, the particular sensitivities inherent to the optimization problem are reflected by the error estimator.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

We will compare the performance of the weighted error estimator (12.7) with a more traditional error indicator. Control of the error in the 'energy norm' of the Euler–Lagrange equations (12.4) alone leads to the *a posteriori* error indicator

$$\eta_E(u_h) := c_i \left(\sum_{K \in \mathbb{T}_h} h_K^2 \left\{ (\rho_K^u)^2 + (\rho_K^z)^2 + (\rho_K^q)^2 \right\} \right)^{1/2}, \tag{12.8}$$

with the residual terms as defined above. This *ad hoc* criterion aims at satisfying the state equation uniformly with good accuracy. However, this concept seems questionable since it does not take into account the sensitivity of the cost functional with respect to the discretization. Capturing these dependencies is the particular feature of the DWR method.

Numerical results (*from Becker, Kapp and Rannacher* (2000*c*))
We consider the configuration as shown in Figure 12.1 with a T-shaped domain Ω of width one. The control acts along the lower boundary Γ_C, whereas the observations are taken along the (longer) upper boundary Γ_O. The cost functional is chosen as in (12.2) with $u_O \equiv 1$ and $\alpha = 1$, that is, the stabilization term constitutes a part of the cost functional.

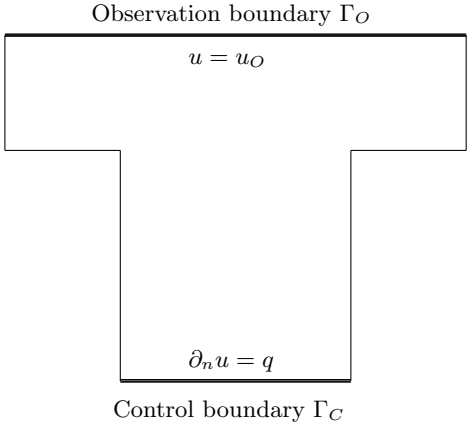
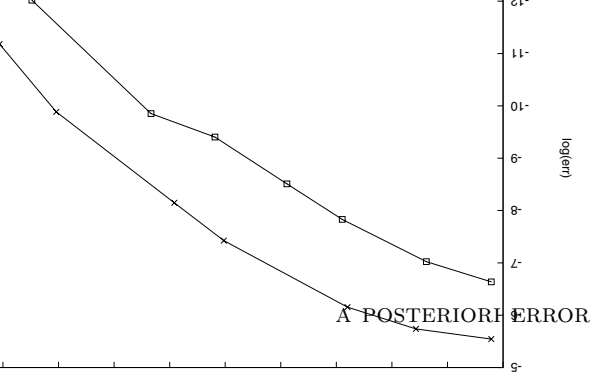


Fig. 12.1. Configuration of the boundary control model problem

Figure 12.2 shows the quality of the weighted error estimator (12.7) for quantitative error control. The relative error $E_{\rm rel}$ and the effectivity index $I_{\rm eff}$ are defined as before. The reference value is obtained on a mesh with more than $200,000$ elements. We compare the weighted error estimator

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press



N	E_{rel}	I_{eff}
320	1.0e-3	1.1
1376	3.5e-4	0.7
4616	3.2e-5	0.7
11816	1.6e-5	1.0
23624	6.4e-6	0.8
48716	2.8e-6	0.7

Fig. 12.2. Effectivity of the weighted error estimator (left), and comparison of the efficiency of the meshes generated by the two estimators, '$\times$' error values by the energy estimator, '$\square$' error values by the weighted estimator (log$-$log scale)

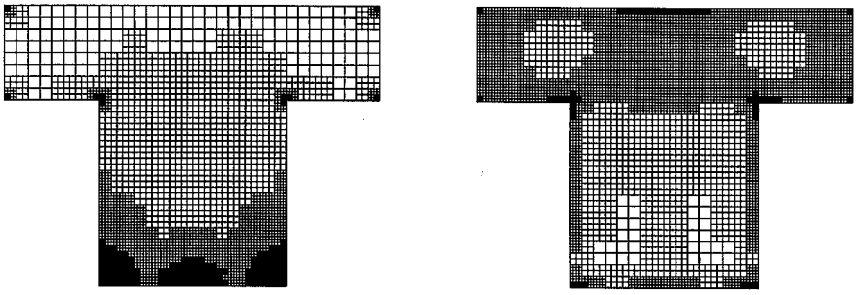


Fig. 12.3. Comparison between meshes obtained by the energy error estimator (left) and the weighted error estimator (right); $N \sim 5,000$ cells in both cases

with the 'energy error' estimator (12.8) for the state equation. Figure 12.3 shows meshes generated by the two estimators.

The difference in the meshes can be explained as follows. Obviously, the energy error estimator observes the irregularities introduced on the control boundary by the jump in the non-homogeneous Neumann condition, but it tends to over-refine in this region and to under-refine at the observation boundary. The weighted error estimator observes the needs of the optimization process by distributing the cells more evenly.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Notes and references

Here we have only presented results for optimal control problems with *linear* state equation. More realistic examples with nonlinear state equation (simplified Ginzburg–Landau model of superconductivity and Navier–Stokes equations in fluid mechanics) have been treated in Becker *et al.* (2000*c*) and Kapp (2000). The minimization of drag in a viscous flow is the subject of Becker (2001, 200x) and Becker, Heuveline and Rannacher (2001*b*).

13. Conclusion and outlook

In this article, we have presented a general approach to error control and mesh adaptation (the DWR method) for finite element approximations of variational problems. By employing optimal-control principles, *a posteriori* error estimates have been derived for the approximation of functionals of the solution related to quantities of physical interest. These error bounds are evaluated by numerically solving linearized 'adjoint problems'. In this context, we have discussed several theoretical and practical aspects of *a posteriori* error estimation and mesh adaptation.

The performance of the DWR method has been demonstrated for several linear and nonlinear model problems mainly from fluid and solid mechanics. In these tests the *dual-weighted* error estimators prove to be asymptotically correct and provide the basis of constructing economical meshes. Effectivity comparisons have been made with some of the traditional heuristic refinement indicators (*e.g.*, ZZ and energy error indicators). The evaluation of the 'weighted' error estimates requires us to solve a linear 'adjoint problem' on each mesh level. In the nonlinear case, this amounts to about the equivalent of *one* extra step within the Newton iteration on this mesh level. In this case, the extra work for mesh adaptation usually makes up 5–25% of the total work on the *optimized* mesh. However, the implementation may appear difficult when existing software components like mesh generators, multigrid solvers, *etc.*, cannot be directly used.

There are several open problems in the theoretical foundation of the DWR method, as well as in its practical realization, that need further investigation.

Theoretical foundation. The strategies for mesh adaptation are largely based on heuristic grounds. One hard open problem is the rigorous proof of the convergence of local residual terms and weights to certain 'limits' for TOL $\to 0$ or $N \to \infty$ as was (heuristically) assumed in (5.5). Further, the extension of the DWR concept for generating solution-adapted *anisotropic* meshes, either by simple cell stretching or by more sophisticated mesh reorientation, is still to be developed.

Control of linearization. In stiff nonlinear problems, *e.g.*, in the neighbourhood of bifurcation points, the linearization in the DWR method

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

involves risks. We have shown that the effect of this linearization can, in principle, be controlled by *a posteriori* strategies. However, the realization of this approach for practical problems is still a critical open problem.

***hp*-finite element method.** In this article the DWR method has been developed only for low-order finite element approximation (piece-wise linear or bilinear shape functions). The extension to higher-order approximation does not pose problems in principle. However, designing effective criteria for simultaneous adaptation of mesh size h and polynomial degree p is still not a fully satisfactorily solved problem, even for 'energy error' control. For the DWR method this problem is largely open.

Time-dependent problems. Multidimensional time-dependent problems constitute a major challenge for the DWR method. Rigorous error control in the space-time frame requires us to solve a space-time adjoint problem. Especially for nonlinear problems, this may be prohibitive with respect to storage space and computing time. The question is how to exploit the option of solving on coarser meshes only and that of data compression. The realization of these concepts for practical problems beyond simple model situations is still in an immature state.

3D problems. The use of the DWR method for spatially 3D problems, though theoretically straightforward, requires special care in setting up the data structure for mesh organization and assembling of the adjoint problem. In particular, the realization for 3D flow problems is a demanding task which has not yet been accomplished.

Finite volume methods. Residual-based methods for *a posteriori* error estimation like the DWR method rely on the variational formulation and the Galerkin orthogonality property of the finite element scheme. This allows us to locally extract additional powers of the mesh size, leading to sensitivity factors of the form $h_K^2\|\nabla^2 z\|_K$. Other 'non-variational' discretizations such as the finite volume methods usually have a different error behaviour, governed by sensitivity factors like $h_K\|\nabla z\|_K$. This may be seen by re-interpreting these discretizations as perturbed finite element schemes obtained by evaluating local integrals by special low-order quadrature rules. It would be desirable to develop a rigorous *a posteriori* error analysis for finite volume schemes following the DWR approach.

Model adaptivity. The concept of *a posteriori* error control for single quantities of interest via duality may also be applicable to other situations when a full model (such as a differential equation) is reduced by projection to a subproblem (such as a finite element model). Model

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

reduction within scales of hierarchical sub-models is a recent development in structural as well as in fluid mechanics (see Babuška and Schwab (1996), Stein and Ohnimus (1997), and Hughes, Feijoo, Mazzei and Quincy (1998)). Quantitative error control in these techniques by computational means like the DWR method seems to be a promising idea.

Adaptivity in optimal control. Most numerical simulation is eventually optimization. Complex multidimensional optimal control and parameter identification problems constitute highly demanding computational tasks. Goal-oriented model reduction by adaptive discretization has high potential to manage large-scale optimization problems in structural and fluid mechanics, such as, for example, minimization of drag, or control of flow-induced structural vibrations. The use of mesh adaptation techniques such as the DWR method for this purpose has just begun.

These are only some of the immediately obvious questions and directions of possible developments. In particular, the practical realization of the DWR method involves many other difficulties varying with the particular features of the problem to be solved, for instance, multi-target mesh adaptation. However, we think the effort invested is worthwhile, even for very complex problems, since, once the method works, the possible gain in accuracy and solution efficiency can be significant.

Acknowledgements

The authors wish to thank several members of the Numerical Analysis Group at their institute for important contributions to this survey article, particularly W. Bangerth, M. Braack, R. Hartmann, G. Kanschat, H. Kapp, F.-T. Suttmeier, C. Waguet, and B. Vexler. They have provided various results for particular cases and much of the computational material. Their comments and proof-reading have greatly helped the presentation.

We emphasize that the whole project would not have been realized without the availability of flexible software for carrying out the numerous calculations involving local mesh adaptation and fast multigrid solution. We credit G. Kanschat, F.-T. Suttmeier and the first author with the development of the finite element C++ package DEAL (Becker, Kanschat and Suttmeier 2001*a*), which in collaboration with W. Bangerth has since grown into the new package DEAL II (Bangerth, Hartmann and Kanschat 2001). This software was used for the computations reported in Sections 3–7 and 10–12. Most of the flow examples in Sections 8 and 9 were computed by the C++ code GASCOIGNE by M. Braack and the first author (Braack and Becker 2001).

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

Last, but not least, we acknowledge the financial support of the Deutsche Forschungsgemeinschaft (DFG) via SFB 359 'Reaktive Strömungen, Diffusion und Transport'.

REFERENCES

M. Ainsworth and J. T. Oden (1993), 'A unified approach to a posteriori error estimation using element residual methods', *Numer. Math.* **65**, 23–50.

M. Ainsworth and J. T. Oden (1997*a*), 'A posteriori error estimation in finite element analysis', *Comput. Meth. Appl. Mech. Eng.* **142**, 1–88.

M. Ainsworth and J. T. Oden (1997*b*), 'A posteriori error estimators for the Stokes and Oseen equations', *SIAM J. Numer. Anal.* **34**, 228–245.

I. Babuška and A. D. Miller (1984*a*), 'The post-processing approach in the finite element method, I: Calculations of displacements, stresses and other higher derivatives of the displacements', *Int. J. Numer. Meth. Eng.* **20**, 1085–1109.

I. Babuška and A. D. Miller (1984*b*), 'The post-processing approach in the finite element method, II: The calculation of stress intensity factors', *Int. J. Numer. Meth. Eng.* **20**, 1111–1129.

I. Babuška and A. D. Miller (1984*c*), 'The post-processing approach in the finite element method, III: A posteriori error estimation and adaptive mesh selection', *Int. J. Numer. Meth. Eng.* **20**, 2311–2324.

I. Babuška and A. D. Miller (1987), 'A feedback finite element method with a posteriori error estimation', *Comput. Meth. Appl. Mech. Eng.* **61**, 1–40.

I. Babuška and W. C. Rheinboldt (1978*a*), 'Error estimates for adaptive finite element computations', *SIAM J. Numer. Anal.* **15**, 736–754.

I. Babuška and W. C. Rheinboldt (1978*b*), 'A posteriori error estimates for the finite element method', *Int. J. Numer. Meth. Eng.* **12**, 1597–1615.

I. Babuška and C. Schwab (1996), 'A posteriori error estimation for hierarchic models of elliptic boundary value problems on thin domains', *SIAM J. Numer. Anal.* **33**, 221–246.

W. Bangerth (1998), 'Finite element approximation of the acoustic wave equation: Error control and mesh adaptation', Diplomarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

W. Bangerth (2000), Mesh adaptivity and error control for a finite element approximation of the elastic wave equation, in *Fifth International Conference on Mathematical and Numerical Aspects of Wave Propagation* (Waves2000) (A. Bermúdez *et al.*, eds), SIAM, pp. 725–729.

W. Bangerth and R. Rannacher (1999), 'Finite element approximation of the acoustic wave equation: Error control and mesh adaptation', *East–West J. Numer. Math.* **7**, 263–282.

W. Bangerth, R. Hartmann and G. Kanschat (2001) deal.II: Differential Equations Analysis Library, Technical Reference. Available from: `http://gaia.iwr.uni-heidelberg.de/~deal/`

R. E. Bank and A. Weiser (1985), 'Some a posteriori error estimators for elliptic partial differential equations', *Math. Comp.* **44**, 283–301.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

R. Becker (1995), 'An adaptive finite element method for the incompressible Navier–Stokes equations on time-dependent domains', Doktorarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

R. Becker (1998*a*), An adaptive finite element method for the Stokes equations including control of the iteration error, in *ENUMATH'97* (H. G. Bock *et al.*, eds), World Scientific, Singapore, pp. 609–620.

R. Becker (1998*b*), Weighted error estimators for finite element approximations of the incompressible Navier–Stokes equations, Technical Report RR-3458, INRIA Sophia-Antipolis.

R. Becker (2001) Adaptive finite elements for optimal control problems, Habilitation thesis, Universität Heidelberg.

R. Becker (200x), 'Mesh adaptation for stationary flow control', *J. Math. Fluid Mech.* To appear.

R. Becker and M. Braack (2000), 'Multigrid techniques for finite elements on locally refined meshes', *Numer. Linear Algebra Appl.* **7**, 363–379.

R. Becker and R. Rannacher (1996*a*), Weighted a posteriori error control in FE methods. Lecture at ENUMATH-95, Paris, September 18–22, 1995, Preprint 96-01, SFB 359, Universität Heidelberg. In *Proc. ENUMATH'97* (H. G. Bock *et al.*, eds), World Scientific, Singapore (1998), pp. 621–637.

R. Becker and R. Rannacher (1996*b*), 'A feed-back approach to error control in finite element methods: Basic analysis and examples', *East–West J. Numer. Math.* **4**, 237–264.

R. Becker, C. Johnson and R. Rannacher (1995), 'Adaptive error control for multigrid finite element methods', *Computing* **55**, 271–288.

R. Becker, M. Braack, R. Rannacher and C. Waguet (1999), 'Fast and reliable solution of the Navier–Stokes equations including chemistry', *Comput. Visual. Sci.* **2**, 107–122.

R. Becker, M. Braack and R. Rannacher (2000*a*), Adaptive finite element methods for flow problems, in *FoCM'99* (A. Iserles, ed.), Cambridge University Press.

R. Becker, H. Kapp and R. Rannacher (2000*b*), Adaptive finite element methods for optimization problems, in *Numerical Analysis 1999* (D. F. Griffiths and G. A. Watson, eds), Chapman and Hall/CRC, London, pp. 21–42.

R. Becker, H. Kapp and R. Rannacher (2000*c*), 'Adaptive finite element methods for optimal control of partial differential equations: Basic concepts', *SIAM J. Control Optimiz.* **39**, 113–132.

R. Becker, G. Kanschat and F.-T. Suttmeier (2001*a*) DEAL: Differential Equations Analysis Library. Available from:
`www.mathematik.uni-dortmund.de/user/lsx/suttmeier/deal/deal.html`

R. Becker, V. Heuveline and R. Rannacher (2001*b*) An optimal control approach to adaptivity in computational fluid mechanics (or the power of elementary calculus), in *Proc. ICFD 2001, March 26–29, 2001, Oxford, UK*, Oxford University Press, to appear.

C. Bernardi, O. Bonnon, C. Langouët and B. Métivet (1995), Residual error indicators for linear problems: Extension to the Navier–Stokes equations, in 9*th Int. Conf. Finite Elements in Fluids*.

H. Blum and F.-T. Suttmeier (1999), 'An adaptive finite element discretisation for a simplified Signorini problem', *Calcolo* **37**, 65–77.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

H. Blum and F.-T. Suttmeier (2000), 'Weighted error estimates for finite element solutions of variational inequalities', *Computing* **65** 119–134.

K. Böttcher and R. Rannacher (1996), Adaptive error control in solving ordinary differential equations by the discontinuous Galerkin method, Technical Report Preprint 96-53, SFB 359, Universität Heidelberg.

M. Braack (1998), 'An adaptive finite element method for reactive flow problems', Doktorarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

M. Braack and R. Becker (2001) GASCOIGNE: Simulation tool for flows including chemical reactions. Available from: `http://gascoigne.uni-hd.de/`

M. Braack and R. Rannacher (1999), Adaptive finite element methods for low-Mach-number flows with chemical reactions, in 30*th Computational Fluid Dynamics* (H. Deconinck, ed.), Vol. 1999-03 of *Lecture Series*, von Karman Institute for Fluid Dynamics.

S. Brenner and R. L. Scott (1994), *The Mathematical Theory of Finite Element Methods*, Springer, Berlin/Heidelberg/New York.

G. Caloz and J. Rappaz (1997), Numerical analysis for nonlinear and bifurcation problems, in *Handbook of Numerical Analysis*, Vol. 5, *Techniques of Scientific Computing*, Part 2 (P. G. Ciarlet and J. L. Lions, eds), North-Holland, Amsterdam.

G. F. Carey and J. T. Oden (1984), *Finite Elements, Computational Aspects*, Vol. III, Prentice-Hall.

C. Carstensen and R. Verfürth (1999), 'Edge residuals dominate a posteriori error estimators for low-order finite element methods', *SIAM J. Numer. Anal.* **36**, 1571–1587.

P. G. Ciarlet (1978), *Finite Element Methods for Elliptic Problems*, North-Holland, Amsterdam.

W. Dörfler and M. Rumpf (1998), 'An adaptive strategy for elliptic problems including a posteriori controlled boundary approximation', *Math. Comp.* **224**, 1361–1382.

K. Eriksson and C. Johnson (1988), 'An adaptive finite element method for linear elliptic problems', *Math. Comp.* **50**, 361–383.

K. Eriksson and C. Johnson (1991), 'Adaptive finite element methods for parabolic problems, I: A linear model problem', *SIAM J. Numer. Anal.* **28**, 43–77.

K. Eriksson and C. Johnson (1993), 'Adaptive streamline diffusion finite element methods for stationary convection-diffusion problems', *Math. Comp.* **60**, 167–188.

K. Eriksson and C. Johnson (1995*a*), 'Adaptive finite element methods for parabolic problems, II: Optimal error estimates in $l_\infty l_2$ and $l_\infty l_\infty$', *SIAM J. Numer. Anal.* **32**, 706–740.

K. Eriksson and C. Johnson (1995*b*), 'Adaptive finite element methods for parabolic problems, IV: Nonlinear problems', *SIAM J. Numer. Anal.* **32**, 1729–1749.

K. Eriksson and C. Johnson (1995*c*), 'Adaptive finite element methods for parabolic problems, V: Long-time integration', *SIAM J. Numer. Anal.* **32**, 1750–1763.

K. Eriksson, D. Estep, P. Hansbo and C. Johnson (1995), Introduction to adaptive methods for differential equations, in *Acta Numerica* (A. Iserles, ed.), Vol. 4, Cambridge University Press, pp. 105–158.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

K. Eriksson, C. Johnson and S. Larsson (1998), 'Adaptive finite element methods for parabolic problems, VI: Analytic semigroups', *SIAM J. Numer. Anal.* **35**, 1315–1325.

D. Estep (1995), 'A posteriori error bounds and global error control for approximation of ordinary differential equations', *SIAM J. Numer. Anal.* **32**, 1–48.

D. Estep and D. French (1994), 'Global error control for the continuous Galerkin finite element method for ordinary differential equations', *Modél. Math. Anal. Numér.* **28**, 815–852.

D. Estep and S. Larsson (1993), 'The discontinuous Galerkin method for semilinear parabolic problems', *Modél. Math. Anal. Numér.* **27**, 611–643.

C. Führer (1996), 'Error control in finite element methods for hyperbolic problems', Doktorarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

C. Führer and G. Kanschat (1997), 'A posteriori error control in radiative transfer', *Computing* **58**, 317–334.

C. Führer and R. Rannacher (1997), 'An adaptive streamline-diffusion finite element method for hyperbolic conservation laws', *East–West J. Numer. Math.* **5**, 145–162.

M. Giles, M. Larsson, M. Levenstam and E. Süli (1997), 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.

P. Hansbo (200x), Three lectures on error estimation and adaptivity, in *Adaptive Finite Elements in Linear and Nonlinear Solid and Structural Mechanics* (E. Stein, ed.), Vol. 416 of *CISM Courses and Lectures*, Springer. To appear.

P. Hansbo and C. Johnson (1991), 'Adaptive streamline diffusion finite element methods for compressible flow using conservative variables', *Comput. Meth. Appl. Mech. Eng.* **87**, 267–280.

R. Hartmann (1998), 'A posteriori Fehlerschätzung und adaptive Schrittweiten- und Ortsgittersteuerung bei Galerkin-Verfahren für die Wärmeleitungsgleichung', Diplomarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

R. Hartmann (2000), A comparison of adaptive SDFEM and DG(r) method for linear transport problems, Technical report, Institut für Angewandte Mathematik, Universität Heidelberg.

R. Hartmann (200x), Adaptive FE-methods for conservation equations, in *8th International Conference on Hyperbolic Problems: Theory, Numerics, Applications* (HYP2000) (G. Warnecke, ed.) To appear.

F.-K. Hebeker and R. Rannacher (1999), 'An adaptive finite element method for unsteady convection-dominated flows with stiff source terms', *SIAM J. Sci. Comput.* **21**, 799–818.

V. Heuveline and R. Rannacher (2001*a*) A posteriori error control for finite element approximations of elliptic eigenvalue problems, Preprint 2001-08 (SFB 359), Universität Heidelberg. To appear in *Adv. Comput. Math.*

V. Heuveline and R. Rannacher (2001*b*) Adaptive finite element discretization of eigenvalue problems in hydrodynamic stability theory, Preprint, Institut für Angewandte Mathematik, Universität Heidelberg.

P. Houston, R. Rannacher and E. Süli (2000), 'A posteriori error analysis for stabilized finite element approximation of transport problems', *Comput. Meth. Appl. Mech. Eng.* **190**, 1483–1508.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

T. J. R. Hughes and A. N. Brooks (1982), 'Streamline upwind/Petrov Galerkin formulations for convection dominated flows with particular emphasis on the incompressible Navier–Stokes equation', *Comput. Meth. Appl. Mech. Eng.* **32**, 199–259.

T. J. R. Hughes, L. P. Franca and M. Balestra (1986), 'A new finite element formulation for computational fluid dynamics, V: Circumvent the Babuška–Brezzi condition: A stable Petrov–Galerkin formulation for the Stokes problem accommodating equal order interpolation', *Comput. Meth. Appl. Mech. Eng.* **59**, 89–99.

T. J. R. Hughes, G. R. Feijoo, L. Mazzei and J.-B. Quincy (1998), 'The variational multiscale method: a paradigm for computational mechanics', *Comput. Meth. Appl. Mech. Eng.* **166**, 3–24.

C. Johnson (1987), *Numerical Solution of Partial Differential Equations by the Finite Element Method*, Cambridge University Press, Cambridge.

C. Johnson (1990), 'Adaptive finite element methods for diffusion and convection problems', *Comput. Meth. Appl. Mech. Eng.* **82**, 301–322.

C. Johnson (1992), 'Adaptive finite element methods for the obstacle problem', *Math. Models Meth. Appl. Sci.* **2**, 483–487.

C. Johnson (1993), 'Discontinuous Galerkin finite element methods for second order hyperbolic problems', *Comput. Meth. Appl. Mech. Eng.* **107**, 117–129.

C. Johnson (1994), A new paradigm for adaptive finite element methods, in *Proc. MAFELAP 93* (J. Whiteman, ed.), Wiley, pp. 105–120.

C. Johnson and P. Hansbo (1992*a*), Adaptive finite element methods for small strain elasto-plasticity, in *Finite Inelastic Deformations: Theory and Applications* (D. Besdo and E. Stein, eds), Springer, Berlin, pp. 273–288.

C. Johnson and P. Hansbo (1992*b*), 'Adaptive finite element methods in computational mechanics', *Comput. Meth. Appl. Mech. Eng.* **101**, 143–181.

C. Johnson and R. Rannacher (1994), On error control in CFD, in *Int. Workshop Numerical Methods for the Navier–Stokes Equations* (F.-K. Hebeker *et al.*, eds), Vol. 47 of *Notes Numer. Fluid Mech*, Vieweg, Braunschweig, pp. 133–144.

C. Johnson and A. Szepessy (1995), 'Adaptive finite element methods for conservation laws based on a posteriori error estimates', *Comm. Pure Appl. Math.* **48**, 199–234.

C. Johnson, R. Rannacher and M. Boman (1995), 'Numerics and hydrodynamic stability: Towards error control in CFD', *SIAM J. Numer. Anal.* **32**, 1058–1079.

G. Kanschat (1996), 'Parallel and adaptive Galerkin methods for radiative transfer problems', Doktorarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

G. Kanschat (1998), 'A robust finite element discretization for radiative transfer problems with scattering', *East–West J. Numer. Math.* **6**, 265–272.

G. Kanschat (2000), Solution of multi-dimensional radiative transfer problems on parallel computers, in *Parallel Solution of Partial Differential Equations* (P. Bjørstad and M. Luskin, eds), Vol. 120 of *IMA Volumes in Mathematics and its Applications*, Springer, New York, pp. 85–96.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

H. Kapp (2000), 'Adaptive finite element methods for optimization in partial differential equations', Doktorarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

P. Ladeveze and D. Leguillon (1983), 'Error estimate procedure in the finite element method and applications', *SIAM J. Numer. Anal.* **20**, 485–509.

J. Lang (1999), 'Adaptive multilevel solution of nonlinear parabolic PDE systems: Theory, algorithms, and applications', Habilitationsschrift, Konrad-Zuse-Zentrum für Informationstechnik, Berlin.

M. G. Larson (2000), 'A posteriori and a priori error estimates for finite element approximations of selfadjoint eigenvalue problems', *SIAM J. Numer. Anal.* **38** 608–625.

M. G. Larson and A. J. Niklasson (1999), 'Adaptive multilevel finite element approximations of semilinear elliptic boundary value problems', *Numer. Math.* **84**, 249–274.

L. Machiels, A. T. Patera and J. Peraire (1998), Output bound approximation for partial differential equations: Application to the incompressible Navier–Stokes equations, in *Industrial and Environmental Applications of Direct and Large Eddy Numerical Simulation* (S. Biringen, H. Örs, A. Tezel and J. H. Ferziger, eds), Springer, Berlin/Heidelberg/New York, pp. 93–108.

L. Machiels, J. Peraire and A. T. Patera (1999), A posteriori finite element output bounds for the incompressible Navier–Stokes equations: Application to a natural convection problem, Technical Report 99-4, MIT FML.

J. Medina, M. Picasso and J. Rappaz (1996), 'Error estimates and adaptive finite elements for nonlinear diffusion–convection problems', *Math. Models Meth. Appl. Sci.* **5**, 689–712.

C. Nystedt (1995), A priori and a posteriori error estimates and adaptive finite element methods for a model eigenvalue problem, Technical Report Preprint NO 1995-05, Department of Mathematics, Chalmers University of Technology.

J. T. Oden and S. Prudhomme (1999), 'On goal-oriented error estimation for elliptic problems: Application to the control of pointwise errors', *Comput. Meth. Appl. Mech. Eng.* **176**, 313–331.

J. T. Oden, W. Wu and M. Ainsworth (1993), 'An a posteriori error estimate for finite element approximations of the Navier–Stokes equations', *Comput. Meth. Appl. Mech. Eng.* **111**, 185–202.

M. Paraschivoiu and A. T. Patera (1998), 'Hierarchical duality approach to bounds for the outputs of partial differential equations', *Comput. Meth. Appl. Mech. Eng.* **158**, 389–407.

J. Peraire and A. T. Patera (1998), Bounds for linear-functional outputs of coercive partial differential equations: Local indicators and adaptive refinement, in *Advances in Adaptive Computational Methods in Mechanics* (P. Ladeveze and J. T. Oden, eds), Elsevier, Amsterdam, pp. 199–215.

P. Le Quere and H. Paillere (2000), Modelling and simulation of natural convection flows with large temperature differences: A benchmark problem for low Mach number solvers. Available from:
`http://m17.limsi.fr/Individu/plq/nonbou/bench_english.html`

R. Rannacher (1998*a*), Error control in finite element computations, in *Proc. Sum-*

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

mer School Error Control and Adaptivity in Scientific Computing (H. Bulgak and C. Zenger, eds), Kluwer Academic Publishers, pp. 247–278.

R. Rannacher (1998*b*), 'A posteriori error estimation in least-squares stabilized finite element schemes', *Comput. Meth. Appl. Mech. Eng.* **166**, 99–114.

R. Rannacher (200x), Duality techniques for error estimation and mesh adaptation in finite element methods, in *Adaptive Finite Elements in Linear and Nonlinear Solid and Structural Mechanics* (E. Stein, ed.), Vol. 416 of *CISM Courses and Lectures*, Springer. To appear.

R. Rannacher and L. R. Scott (1982), 'Some optimal error estimates for piecewise linear finite element approximations', *Math. Comp.* **38**, 437–445.

R. Rannacher and F.-T. Suttmeier (1997), 'A feed-back approach to error control in finite element methods: Application to linear elasticity', *Comput. Mechanics* **19**, 434–446.

R. Rannacher and F.-T. Suttmeier (1998), 'A posteriori error control in finite element methods via duality techniques: Application to perfect plasticity', *Comput. Mechanics* **21**, 123–133.

R. Rannacher and F.-T. Suttmeier (1999), 'A posteriori error estimation and mesh adaptation for finite element models in elasto-plasticity', *Comput. Meth. Appl. Mech. Eng.* **176**, 333–361.

R. Rannacher and F.-T. Suttmeier (200x), Error estimation and adaptive mesh design for FE models in elasto-plasticity, in *Error-Controlled Adaptive FEMs in Solid Mechanics* (E. Stein, ed.), Wiley. To appear.

M. Schäfer and S. Turek (1996), Benchmark computations of laminar flow around a cylinder (with support by F. Durst, E. Krause and R. Rannacher), in *Flow Simulation with High-Performance Computers, II: DFG Priority Research Program Results 1993–1995* (E. H. Hirschel, ed.), Vol. 52 of *Notes Numer. Fluid Mech.*, Vieweg, Wiesbaden, pp. 547–566.

J. Segatz, R. Rannacher, J. Wichmann, C. Orlemann, T. Dreier and J. Wolfrum (1996), 'Detailed numerical simulation in flow reactors: A new approach in measuring absolute rate constants', *J. Phys. Chem.* **100**, 9323–9333.

E. Stein and S. Ohnimus (1997), 'Coupled model- and solution-adaptivity in the finite-element method', *Comput. Meth. Appl. Mech. Eng.* **150**, 327–350.

F.-T. Suttmeier (1996), 'Adaptive finite element approximation of problems in elasto-plasticity theory', Doktorarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

F.-T. Suttmeier (1998), 'An adaptive displacement/pressure finite element scheme for treating incompressibility effects in elasto-plastic materials'. Preprint 1998-33, SFB 359, Universität Heidelberg. To appear in *Numer. Meth. Part. Diff. Equ.*

F.-T. Suttmeier (2001) General approach for a posteriori error estimates for finite element solutions of variational inequalities, *Comp. Mech.* **27**, 317–323.

R. Verfürth (1993), A posteriori error estimates for nonlinear problems, in *Numerical Methods for the Navier–Stokes Equations* (F.-K. Hebeker *et al.*, eds), Vol. 47 of *Notes Numer. Fluid Mech.*, Vieweg, Braunschweig, pp. 288–297.

R. Verfürth (1994), 'A posteriori error estimates for nonlinear problems: Finite element discretization of elliptic equations', *Math. Comp.* **62**, 445–475.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press

R. Verfürth (1996), *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley/Teubner, New York/Stuttgart.

R. Verfürth (2000), 'A posteriori error estimation techniques for nonlinear elliptic and parabolic pde's', *Rev. Eur. Elem. Finis* **9**, 377–402.

B. Vexler (2000), 'A posteriori Fehlerschätzung und Gitteradaption bei Finite-Elemente-Approximationen nichtlinearer elliptischer Differentialgleichungen', Diplomarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

C. Waguet (2000), 'Adaptive finite element computation of chemical flow reactors', Doktorarbeit, Institut für Angewandte Mathematik, Universität Heidelberg.

R. Wehrse, E. Meinkoehn and G. Kanschat (1999), Recent work in Heidelberg on the solution of the radiative transfer equation, in *Forum du GRETA, Transfert de Rayonnement en Astrophysique* (P. Stee, ed.), CNRS-INSU-ASPS, pp. 82–99.

O. C. Zienkiewicz and J. Z. Zhu (1987), 'A simple error estimator and adaptive procedure for practical engineering analysis', *Int. J. Numer. Meth. Eng.* **24**, 337–357.

S. Ziukas and N.-E. Wiberg (1998), Adaptive procedure with superconvergent patch recovery for linear parabolic problems, in *Finite Element Methods: Superconvergence, Post-Processing, and A Posteriori Estimates* (M. Krizek *et al.*, eds), Vol. 196 of *Inc. Lect. Notes Pure Appl. Math.*, Marcel Dekker, pp. 303–314.

https://doi.org/10.1017/S0962492901000010 Published online by Cambridge University Press