CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

(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

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

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 $J(\cdot)$.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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.

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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

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

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

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

© Cambridge University Press www.cambridge.org

CAMBRIDGE

Cambridge University Press
978-0-521-80312-0 - Acta Numerica 2001, Volume 10
Edited by Arieh Iserles
Excerpt
More information

'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

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

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,

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

© Cambridge University Press www.cambridge.org