Chapter 6
Balancing Discretization and Iteration Error in Finite Element A Posteriori Error Analysis

Rolf Rannacher and Jevgeni Vihharev

Abstract This article surveys recent developments in a combined a posteriori analysis for the discretization and iteration errors in the finite element approximation of elliptic PDE systems. The underlying theoretical framework is that of the **D**ual **W**eighted **R**esidual (DWR) method for goal-oriented error control. Based on computable a posteriori error estimates the algebraic iteration can be adjusted to the discretization in a successive mesh adaptation process. The performance of the proposed method is demonstrated for several model situations including the simple Poisson equation, the Stokes equations in fluid mechanics and the KKT system of a linear-quadratic elliptic optimal control problem. Furthermore, extensions are discussed for certain classes of nonlinear problems including eigenvalue problems and nonlinear reaction-diffusion equations.

6.1 Introduction

The use of adaptive techniques based on a posteriori estimates for the discretization error is well accepted in the context of finite element discretization of partial differential equations (see, e.g., [1, 8, 25]). Although the convergence properties of linear as well as nonlinear iterative methods such as the multigrid method or the Newton method are discussed in many publications (see, e.g., [3, 10–12, 14]), there are only few results on a posteriori error estimation of the iteration error. In the case of solving the Poisson equation, work has been done in [6] and was extended to the Stokes equations in [4]. There, the automatic control of the discretization and multigrid errors has been developed with respect to L^2- and energy norms. The reliability of the proposed adaptive algorithm has been verified on uniformly refined meshes.

However, in many applications, the error measured in global norms does not provide useful bounds for the error in terms of a given functional, a so-called *quantity*

R. Rannacher (✉) · J. Vihharev
Institute of Applied Mathematics, University of Heidelberg, INF 293/294, 69120 Heidelberg, Germany
e-mail: rolf.rannacher@iwr.uni-heidelberg.de

J. Vihharev
e-mail: jevgeni.vihharev@iwr.uni-heidelberg.de

S. Repin et al. (eds.), *Numerical Methods for Differential Equations, Optimization, and Technological Problems*, Computational Methods in Applied Sciences 27,
DOI 10.1007/978-94-007-5288-7_6,

of interest. In this work, we propose the simultaneous control of both discretization and iteration errors with respect to a prescribed output functional. This approach is based on a posteriori error estimation by dual weighted residuals as presented in [8] as the **D**ual **W**eighted **R**esidual (DWR) method. We incorporate the adaptive iteration method into the solution process of a given problem. It seems natural to stop the linear or nonlinear iteration when the error due to the approximate solution of the discrete equations is comparable to the error due to the finite element discretization itself. To this purpose, we derive an a posteriori error estimator which simultaneously assesses the influences of the discretization and the inexact solution of the arising algebraic equations. This allows us to balance both sources of errors.

For illustration, we consider the model problem

$$Au = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \Gamma, \tag{6.1}$$

with a linear elliptic operator A and a right-hand side $f \in L^2(\Omega)$ where Ω is assumed to be a bounded domain in $\mathbb{R}^d$, $d \in \{2, 3\}$, with polygonal respectively polyhedral boundary Γ. For simplicity, we impose homogeneous Dirichlet boundary conditions. However, the techniques developed in this paper can also be applied to problems with other types of boundary conditions. For the variational formulation of the problem (6.1), we introduce the Hilbert space $V := H_0^1(\Omega)$ and the L^2-scalar product $(v, w) := (v, w)_{L^2(\Omega)}$. With the bilinear form $a(\cdot, \cdot) \colon V \times V \to \mathbb{R}$ associated to the linear operator A, the weak formulation of the problem (6.1) reads as follows: Find $u \in V$ such that

$$a(u, \phi) = (f, \phi) \quad \forall \phi \in V. \tag{6.2}$$

We discretize this problem by a standard finite element method (see [13]) in finite dimensional spaces $V_h \subset V$ resulting in "discrete" problems

$$a(u_h, \phi_h) = (f, \phi_h) \quad \forall \phi_h \in V_h, \tag{6.3}$$

which are equivalent to linear systems of algebraic equations. Usually the a posteriori error estimators for the discretization error $u - u_h$ are derived under the assumption that the discrete problems (6.3) are solved exactly. This ensures the crucial property of the Galerkin orthogonality,

$$a(u - u_h, \phi_h) = 0, \quad \phi_h \in V_h. \tag{6.4}$$

In contrast, here, we assume that the discrete problems are solved only approximately and denote the obtained approximate solution in V_h by $\tilde{u}_h$ in contrast to the notation u_h for the "exact" discrete solution. Let the quantity of interest $J(u)$ of the computation be given in terms of a linear functional $J \colon V \to \mathbb{R}$. Our goal is the derivation of an a posteriori error estimate of the form

$$\left| J(u) - J(\tilde{u}_h) \right| \le \eta_h + \eta_{\text{it}}. \tag{6.5}$$

Here, η_h and η_{it} denote error estimators which can be evaluated from the computed discrete solution $\tilde{u}_h$, where η_h assesses the error due to the finite element

discretization and η_{it} the error due to the inexact solution of the discrete equations. The adaptation strategy then aims at equilibrating these two error components, $\eta_{\text{it}} \approx \eta_h \approx \frac{1}{2}\text{TOL}$, according to the prescribed error tolerance TOL. This results in a practical stopping criterion for the linear or nonlinear algebraic iteration.

This article is based on the results of the articles [6, 17, 20, 21]. The outline is as follows: In Sect. 6.2, we describe the finite element discretization of the problem (6.1) and develop the principles of the DWR method for goal-oriented a posteriori error estimation of the discretization as well as the iteration errors. Section 6.2.1 discusses the practical evaluation of these error estimators and the implementation of the resulting adaptation strategies. The numerical results presented in Sect. 6.2.2 demonstrate the efficiency and reliability of the proposed method for a prototypical scalar model problem. In Sect. 6.3 this approach is developed for the associated symmetric eigenvalue problem. Then, Sect. 6.4 is devoted to the treatment of different types of saddle point problems, the Stokes system in fluid mechanics, and the Karush-Kuhn-Tucker (KKT) system in linear-quadratic optimization. Finally, in Sect. 6.5, we consider the extension of our theory to the Newton iteration for nonlinear elliptic problems. The article concludes with Sect. 6.6, which addresses current work and open problems.

6.2 Goal-Oriented Mesh Adaptation: The DWR Method

We briefly sketch the essentials of "goal-oriented" a posteriori error estimation and mesh adaptation underlying the **D**ual **W**eighted **R**esidual (DWR) method [2, 7, 8].

Let the goal of the computation be the approximation of a scalar quantity $J(u)$ with maximal accuracy TOL on a mesh $\mathbb{T}_h$ from models

$$\mathscr{A}(u) = 0, \qquad \mathscr{A}_h(u_h) = 0.$$

In this process the goal of adaptivity is the optimization of the mesh $\mathbb{T}_h$ guided by an a posteriori error estimate of the form

$$J(u) - J(u_h) \approx \eta(u_h) := \sum_{K \in \mathbb{T}_h} \rho_K(u_h)\omega_K$$

with local cell residuals $\rho_K(u_h)$ and weights ω_K (sensitivity factors). Then, the mesh adaptation is driven by the local error indicators $\eta_K := \rho_K(u_h)\omega_K$. The inherent problem in this approach is that usually only an approximation $\tilde{u}_h$ of the exact discrete solution u_h is available obtained by a nonlinear or linear iteration process.

For illustration, we consider the following model situation. For the solution of the boundary value problem

$$-\Delta u = f \quad \text{in } \Omega \subset \mathbb{R}^2,\ u_{|\partial\Omega} = 0, \tag{6.6}$$

the quantity $J(u)$ is to be determined, where $J(\cdot)$ is a linear functional defined on the natural solution space of this problem. The variational formulation of (6.6) reads

$$u \in V: \quad a(u, \psi) := (\nabla u, \nabla \psi) = (f, \psi) \quad \forall \psi \in V, \tag{6.7}$$

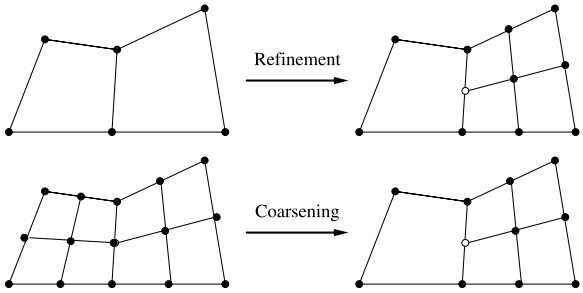

Fig. 6.1 Mesh refinement and coarsening using "hanging nodes"

where $V := H_0^1(\Omega)$ is the usual first-order Sobolev Hilbert space. For approximating this variational problem, we consider a Galerkin finite element method using subspaces $V_h \subset V$ (P_1 or Q_1 elements):

$$u_h \in V_h: \quad a(u_h, \psi_h) = (f, \psi_h) \quad \forall \psi_h \in V_h. \tag{6.8}$$

The spaces V_h are defined on form-regular decompositions $\mathbb{T}_h = \{K\}$ of $\overline{\Omega}$ consisting of closed cells K (triangular/quadrilateral in 2D and tetrahedral/hexahedral in 3D) with diameter h_K (see [13]). The global mesh size is $h := \max_{K\in\mathbb{T}_h} h_K$. To ease local mesh adaptation, we allow "hanging nodes" (at most one per face or edge) where the corresponding "irregular" nodal values are eliminated from the system by linear interpolation of neighboring regular nodal values (see Fig. 6.1).

The error $e := u - u_h$ satisfies the Galerkin orthogonality relation

$$a(e, \psi_h) = 0, \quad \psi_h \in V_h. \tag{6.9}$$

We introduce the associated continuous and discrete "dual" problems

$$z \in V: \quad a(\phi, z) = J(\phi) \quad \forall \phi \in V, \tag{6.10}$$

$$z_h \in V_h: \quad a(\phi_h, z_h) = J(\phi_h) \quad \forall \phi_h \in V_h. \tag{6.11}$$

Taking the test function $\phi = e$ in (6.10) yields the error identity

$$J(e) = a(e, z) = a(e, z - \psi_h) = (f, z - \psi_h) - a(u_h, z - \psi_h) =: \rho(u_h)(z - \psi_h)$$

with an arbitrary $\psi_h \in V_h$. By cell-wise integration by parts, we obtain

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

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

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

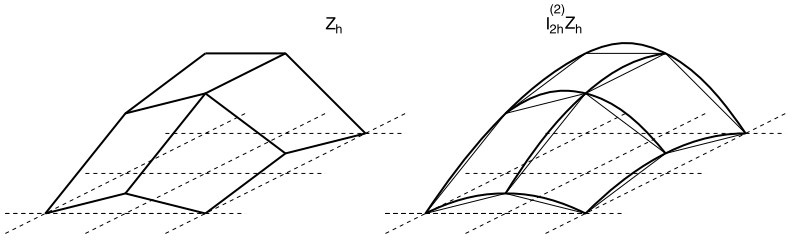

Fig. 6.2 Local post-processing by higher-order patchwise interpolation: "biquadratic" interpolation of computed "bilinear" nodal values

where $[\nabla u_h]$ denotes the jump of the normal derivative across interelement edges. Then, using the refinement indicators

$$\eta_K := \left| \big(R(u_h), z - \psi_h\big)_K + \big(r(u_h), z - \psi_h\big)_{\partial K} \right|,$$

the mesh adaptation aims at "error balancing", i.e.,

$$\eta := \sum_{K \in \mathbb{T}_h} \eta_K, \qquad N := \#\{K \in \mathbb{T}_h\}, \quad \eta_K \approx \mathrm{TOL}/N,$$

which at the end results in $\eta \approx \mathrm{TOL}$.

The unknown dual solution z occurring in the error indicators η_K is approximated by local higher-order post-processing from the computed dual solution z_h,

$$z - I_h^{(1)} z \;\approx\; I_{2h}^{(2)} z_h - z_h,$$

where $I_h^{(1)}$ and $I_{2h}^{(2)}$ denote the operators of cell-wise bilinear and patch-wise biquadratic interpolation, respectively (see Fig. 6.2). This results in the approximate error estimator

$$\begin{aligned} \big|J(e)\big| &\approx \sum_{K \in \mathbb{T}_h} \tilde{\eta}_K, \\ \tilde{\eta}_K &:= \left| \big(R(u_h), I_{2h}^{(2)} z_h - z_h\big)_K + \big(r(u_h), I_{2h}^{(2)} z_h - z_h\big)_{\partial K} \right|. \end{aligned} \tag{6.12}$$

This is to be compared with the traditional global "energy-norm" error estimator

$$\big\|\nabla(u - u_h)\big\| \le \eta_E := c_I c_S \left(\sum_{K \in \mathbb{T}_h} h_K^2 \rho_K(u_h)^2 \right)^{1/2} \tag{6.13}$$

with the cell residuals

$$\rho_K(u_h) := \big(\big\| R(u_h) \big\|_K^2 + \tfrac{1}{2} \big\| r(u_h) \big\|_{\partial K}^2 \big)^{1/2}$$

and certain interpolation and stability constants $c_I \approx 1$ and $c_S \approx 1$.

6.2.1 Balancing of Iteration and Discretization Error

In practice, the "exact" discrete solution $u_h \in V_h$ on the current mesh $\mathbb{T}_h$ is not known but rather an approximation $\tilde{u}_h \in V_h$ obtained by an iterative process $u_h^k \to u_h$ $(k \to \infty)$, such as a simple fixed point method (Gauß-Seidel), a Krylov space method (PCG), or a multigrid method (MG). Hence, in the a posteriori error representation

$$J(e) = \eta := \rho(u_h)(z - \psi_h),$$

we have to use this approximation $\tilde{u}_h := u_h^k$,

$$J(\tilde{e}) \approx \tilde{\eta} := \rho(\tilde{u}_h)(z - \psi_h) + ?$$

We need to balance the "iteration error" and the "discretization error" in order to have a useful stopping criterion (or fine tuning) for the iteration. Suppose that the adaptation process has generated a successively refined sequence of meshes $\mathbb{T}_l := \mathbb{T}_{h_l}$, $l = 0, \ldots, L$, and corresponding approximate discrete solutions $u_l \in V_l := V_{h_l}$.

Algorithm 6.1 Multigrid iteration MG($\mathbf{l}, \gamma, \mathbf{u_l^k}, \mathbf{f_l}$)

if $l = 0$ **then**
  Solve $A_0 u_0^{k+1} = f_0$ exactly.
else
  Pre-smoothing: $\bar{u}_l^k := S_l^{\nu}(u_l^k)$
  Residual: $d_l^k := f_l - A_l \bar{u}_l^k$
  Restriction: $\tilde{d}_{l-1}^k := r_l^{l-1} d_l^k$ (L^2 projection)
  for $r = 1$ **to** γ **do**
    Starting with $v_{l-1}^0 := 0$ iterate $v_{l-1}^r := \mathrm{MG}(l-1, \gamma, v_{l-1}^{r-1}, \tilde{d}_{l-1}^k)$
  end for
  Correction: $\bar{\bar{u}}_l^k := \bar{u}_l^k + p_{l-1}^l \tilde{v}_{l-1}^{\gamma}$ (natural embedding)
  Post-smoothing: $u_l^{k+1} := S_l^{\mu}(\bar{\bar{u}}_l^k)$
end if

Theorem 6.1 *Let $\tilde{u}_L, \tilde{z}_L \in V_L$ be any approximations to the exact primal and dual discrete solutions $u_L, z_L \in V_L$, respectively, on the finest mesh $\mathbb{T}_L$. Then, there holds the error representation*

$$J(u - \tilde{u}_L) = \rho(\tilde{u}_L)(z - \hat{z}_L) + \rho(\tilde{u}_L)(\hat{z}_L). \tag{6.14}$$

If a MG method has been used with canonical components, the following refined representation holds:

$$\rho(\tilde{u}_L)(\hat{z}_L) = \sum_{l=1}^{L} \big(R_l(\tilde{v}_l), \hat{z}_l - \hat{z}_{l-1}\big). \tag{6.15}$$

Here, $\hat{z}_l \in V_l$, $l = 0, \ldots, L$, can be chosen arbitrarily and $R_l(\tilde{v}_l)$ are the iteration residuals on the mesh levels $l = 0, \ldots, L$.

Proof [17] For the error $\tilde{e}_L := u - \tilde{u}_L$ there holds

$$\begin{aligned} J(\tilde{e}_L) &= a(\tilde{e}_L, z) = a(\tilde{e}_L, z - \hat{z}_L) + a(\tilde{e}_L, \hat{z}_L) \\ &= (f, z - \hat{z}_L) - a(\tilde{u}_L, z - \hat{z}_L) + (f, \hat{z}_L) - a(\tilde{u}_L, \hat{z}_L) \\ &= \rho(\tilde{u}_L)(z - \hat{z}_L) + \rho(\tilde{u}_L)(\hat{z}_L). \end{aligned}$$

If the multigrid method has been used, then the second term corresponding to the iteration error can be rewritten in the form

$$\rho(\tilde{u}_L)(\hat{z}_L) = \sum_{l=1}^{L} \big\{ (f, \hat{z}_l - \hat{z}_{l-1}) - a(\tilde{u}_L, \hat{z}_l - \hat{z}_{l-1}) \big\} + \big\{ (f, \hat{z}_0) - a(\tilde{u}_L, \hat{z}_0) \big\}.$$

Since $V_l \subset V_L$ for $l \leq L$, we observe by the definitions of Q_l (Ritz projection), P_l (L^2 projection), and A_l (discrete Laplacian) that for $\phi_l \in V_l$ there holds

$$(f, \phi_l) - a(\tilde{u}_L, \phi_l) = (P_l f, \phi_l) - (A_l Q_l \tilde{u}_L, \phi_l).$$

Further, by the identity $A_l Q_l = P_l A_L$ for $l \leq L$, we have

$$(P_l f, \phi_l) - (A_l Q_l \tilde{u}_L, \phi_l) = \big(P_l(f - A_L \tilde{u}_L), \phi_l\big) = \big(R_l(\tilde{u}_L), \phi_l\big).$$

Using the particular structure of the multigrid method, there holds

$$\begin{aligned} R_l(\tilde{u}_L) &= P_l(f_L - A_L \tilde{u}_L) \\ &= P_l f_L - P_l A_L S_L^{\nu}\big(\tilde{u}_L^{(0)}\big) - P_l A_L p_{L-1}^{L} \tilde{v}_{L-1} \\ &= P_l(d_L - A_{L-1} \tilde{v}_{L-1}) \\ &= P_l d_L - P_l A_{L-1} S_{L-1}^{\nu}\big(\tilde{v}_{L-1}^{(0)}\big) - P_l A_{L-1} p_{L-2}^{L-1} \tilde{v}_{L-2} \\ &\;\;\vdots \\ &= P_l(d_{l+2} - A_l \tilde{v}_{l+1}) \\ &= P_l d_{l+2} - P_l A_{l+1} S_{l+1}^{\nu}\big(\tilde{v}_{l+1}^{(0)}\big) - P_l A_{l+1} p_l^{l+1} \tilde{v}_l \\ &= P_l(d_{l+1} - A_l \tilde{v}_l) = R_l(\tilde{v}_l). \end{aligned}$$

Using this for $\phi_l = \hat{z}_l - \hat{z}_{l-1}$ and $\phi_0 = \hat{z}_0$ completes the proof. □

On the basis of the error representation (6.14), we use the following error balancing criterion:

$$\big|\rho(\tilde{u}_L)(\hat{z}_L)\big| \ll \big|\rho(\tilde{u}_L)(z - \hat{z}_L)\big|. \tag{6.16}$$

Since $\rho(u_L)(\hat{z}_L) = 0$ the term on the left tends to zero for proceeding iteration while the term on the right approaches the (generally) non-zero discretization error. Therefore, the left-hand term can be interpreted as measuring deviation from Galerkin

orthogonality of $\tilde{u}_L$ and the right-hand term is used for estimating the discretization error, however evaluated at the approximative solution $\tilde{u}_L$, i.e.,

$$\big|J(u-u_L)\big| \approx \rho(\tilde{u}_L)(z-\hat{z}_L), \qquad \big|J(u_L-\tilde{u}_L)\big| \approx \rho(\tilde{u}_L)(\hat{z}_L). \tag{6.17}$$

This heuristic concept is supported by the results of the test calculations presented below. It seems to be valid even on coarser meshes provided that the algebraic iteration is organized in a nested fashion, i.e., the approximate solution on the mesh $\mathbb{T}_{l-1}$ is used as the starting value for the iteration on the next refined mesh $\mathbb{T}_l$.

Remark 6.1 It is worth noting that:

1. The proof of the analogue of Theorem 6.1 for "energy-norm" and L^2-norm error control is due to [6].
2. The first error representation,

$$J(\tilde{e}_L) = \rho(\tilde{u}_L)(z-\hat{z}_L) + \rho(\tilde{u}_L)(\hat{z}_L),$$

   can be used for approximative solutions $\tilde{u}_L$ obtained by any solution process in V_L, such as simple fixed point iterations, Krylov space methods, or multigrid methods as well as perturbations caused by numerical quadrature or other "variational crimes".
3. The second error representation holds for V-, W-, or F-cycles and for any type of smoothing. It allows not only balancing the iteration against the discretization error but also tuning the smoothing iteration separately on the different mesh levels,

$$J(\tilde{e}_L) = \rho(\tilde{u}_L)(z-\hat{z}_L) + \sum_{l=1}^{L}\big(R_l(\tilde{v}_l), \hat{z}_l - \hat{z}_{l-1}\big).$$

   The corresponding adaptive algorithm is formulated below.

Algorithm 6.2 Adaptive algorithm

Choose an initial discretization $\mathbb{T}_{h_0}$ and set $l = 0$.
loop
  Set $k = 1$
  repeat
    if $k = 1$ **then**
      for $j = 0$ to l **do**
        Set $\nu_j = 1$, $\mu_j = 1$.
      end for
    end if
    Apply one multigrid cycle to the problem $A_l u_l = f_l$.
    Set $k = k + 1$.
    Evaluate the estimators η_{m_l} and η_{h_l}.

According to the error indicators on the different levels, $(R_j(\tilde{v}_j), \tilde{z}_j - p^j_{j-1}\tilde{z}_{j-1})$ determine the subset of levels $I = \{i_1, \ldots, i_n\}$ with the biggest contribution to the error estimator and increase the number of smoothing steps by
if $k > 1$ **then**
for $j = 1$ to n **do**
Set $\nu_{i_j} = 4$, $\mu_{i_j} = 4$.
end for
end if
until $|\eta_{m_l}| \leq c|\eta_{h_l}|$
if $|\eta_{h_l} + \eta_{m_l}| \leq \mathrm{TOL}$ **then**
stop
end if
Refine the mesh $\mathbb{T}_{h_l} \to \mathbb{T}_{h_{l+1}}$ accordingly to size of $\eta_{h_l,i}$.
Interpolate the previous solution $\tilde{u}_l$ on the mesh $\mathbb{T}_{h_{l+1}}$.
Increment l.
end loop

6.2.2 Numerical Tests

We consider a model Poisson problem (6.6) on a L-shaped domain $\Omega \subset \mathbb{R}^2$. The target value is the function value $J(u) := u(a)$ where $a = (0.2, 0.2)$. This irregular functional is regularized by

$$J_\varepsilon(u) := \big|B_\varepsilon(a)\big|^{-1} \int_{B_\varepsilon(a)} u(x)\,dx = u(a) + O\big(\varepsilon^2\big).$$

The discrete problems are solved by an MG method using a V-cycle and $4 + 4$ ILU-smoothing steps. The tolerance is $\mathrm{TOL} = 5 \times 10^{-7}$. By "MG I", we indicate iteration towards a round-off error level, while "MG II" refers to the use of an adaptive stopping criterion. The computational results are shown in Figs. 6.3, 6.4, and 6.5 as well as Tables 6.1 and 6.2. The "effectivity indices" for measuring the quality of the a posteriori error estimators are defined by

$$I^{\mathrm{tot}}_{\mathrm{eff}} := \frac{|J(e)|}{\eta_h + \eta_{\mathrm{it}}}, \qquad I^{h}_{\mathrm{eff}} := \frac{|J(e_h)|}{\eta_h}, \qquad I^{\mathrm{it}}_{\mathrm{eff}} := \frac{|J(e_{\mathrm{it}})|}{\eta_{\mathrm{it}}}.$$

Next, we consider the computation of the approximate solution u_h on a fixed locally refined, but still rather coarse, mesh by the Gauß-Seidel and the conjugate gradient (CG) method. The computational results are shown in Tables 6.3 and 6.4. In all cases the adaptive strategies proposed lead to significant work savings. Furthermore, the effectivity indices are close to one on finer meshes, which confirms the quality of the error estimators.

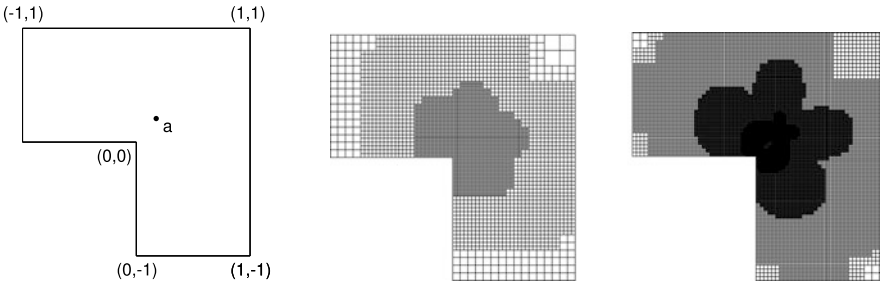

Fig. 6.3 Configuration and locally refined meshes

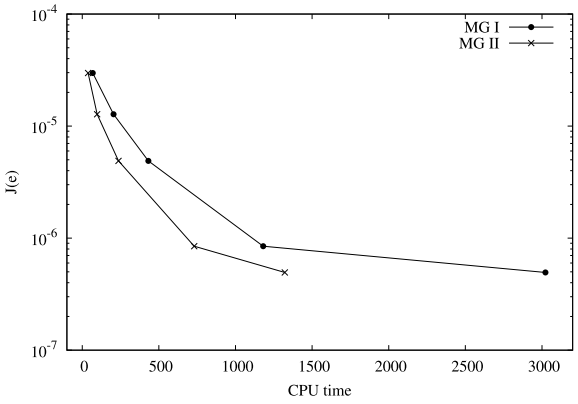

Fig. 6.4 Comparison of the CPU time used by the different MG methods MG I and MG II

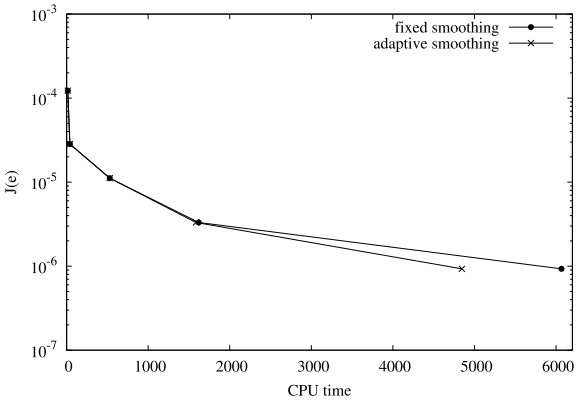

Fig. 6.5 Gain in efficiency of the multigrid algorithm by the adaptive choice of smoothing type and number of steps on the different mesh levels: $1+1$ ILU steps or $4+4$ ILU steps

Table 6.1 Iteration with *MG I* (iteration towards a round-off error level)

N	# Iter	$J(e)$	$\eta_h + \eta_{\text{it}}$	η_h	η_{it}	$I_{\text{eff}}^{\text{tot}}$
225	5	4.06e-03	1.57e-03	1.57e-03	6.01e-14	2.56
721	6	1.16e-03	9.57e-04	9.57e-04	3.95e-14	1.21
1 625	7	4.35e-04	2.26e-04	2.26e-04	4.70e-14	1.92
4 573	8	1.43e-04	9.95e-05	9.95e-05	7.71e-13	1.43
11 565	8	5.50e-05	2.98e-05	2.98e-05	1.67e-12	1.85
31 077	10	1.85e-05	1.28e-05	1.28e-05	6.33e-13	1.43
67 669	9	5.94e-06	4.89e-06	4.89e-06	2.67e-12	1.22
174 585	10	8.47e-07	2.00e-06	2.00e-06	1.79e-12	2.38
427 185	10	4.94e-07	7.63e-07	7.63e-07	1.37e-12	0.64

Table 6.2 Iteration with *MG II* (an adaptive stopping criterion)

N	# Iter	$J(e)$	$\eta_h + \eta_{\text{it}}$	η_h	η_{it}	$I_{\text{eff}}^{\text{tot}}$
225	1	4.06e-03	1.67e-03	1.58e-03	9.42e-05	2.44
721	2	1.16e-03	9.58e-04	9.57e-04	1.35e-06	1.21
1 625	1	4.35e-04	2.44e-04	2.26e-04	1.89e-05	1.19
4 573	2	1.43e-04	1.01e-04	9.95e-05	1.28e-06	1.43
11 565	2	5.50e-05	3.04e-05	2.98e-05	6.43e-07	1.82
31 077	2	1.85e-05	1.40e-05	1.28e-05	1.23e-06	1.32
67 669	2	5.94e-06	5.36e-06	4.89e-06	4.71e-07	1.11
174 585	3	8.47e-07	2.05e-06	2.00e-06	5.04e-08	0.41
427 185	3	4.94e-07	8.04e-07	7.63e-07	4.07e-08	0.64

Table 6.3 Gauss-Seidel iteration on a locally refined mesh with 721 knots (starting value taken from the preceding mesh)

Iter	$J(e_h)$	η_h	I_{eff}^h	$J(e_{\text{it}})$	η_{it}	$I_{\text{eff}}^{\text{it}}$	$\|u_L^{(k)} - u_L\|_\infty$
10	1.16e-3	9.42e-4	1.24	1.68e-3	1.65e-3	1.02	4.21e-2
20	1.16e-3	9.48e-4	1.22	1.21e-3	1.20e-3	1.01	3.66e-2
30	1.16e-3	9.51e-4	1.22	9.10e-4	9.01e-4	1.01	3.20e-2
40	1.16e-3	9.53e-4	1.22	6.86e-4	6.81e-4	1.01	2.78e-2
50	1.16e-3	9.54e-4	1.22	5.18e-4	5.15e-4	1.01	2.42e-2
60	1.16e-3	9.55e-4	1.22	3.90e-4	3.88e-4	1.00	2.10e-2
70	1.16e-3	9.55e-4	1.22	2.94e-4	2.93e-4	1.00	1.83e-2
80	1.16e-3	9.56e-4	1.22	2.21e-4	2.21e-4	1.00	1.59e-2
90	1.16e-3	9.56e-4	1.22	1.67e-4	1.66e-4	1.00	1.38e-2
100	1.16e-3	9.56e-4	1.22	1.25e-4	1.25e-4	1.00	1.19e-2

Table 6.4 CG iteration on a locally refined mesh with 721 knots (the starting value taken from the preceding mesh)

Iter	$J(e_h)$	η_h	I_{eff}^h	$J(e_{\text{it}})$	η_{it}	$I_{\text{eff}}^{\text{it}}$	$\|b - Ax^{(k)}\|_{A^{-1}}$
5	1.16e-3	9.50e-4	1.24	1.85e-03	1.80e-03	1.03	7.57e-3
10	1.16e-3	9.54e-4	1.22	4.60e-04	4.50e-04	1.03	6.34e-3
15	1.16e-3	9.50e-4	1.24	3.10e-05	2.99e-05	1.04	1.17e-3
20	1.16e-3	9.55e-4	1.22	2.17e-05	2.17e-05	1.01	3.08e-4
25	1.16e-3	9.57e-4	1.22	4.12e-06	4.12e-06	1.01	1.01e-4
30	1.16e-3	9.57e-4	1.22	1.09e-06	1.09e-06	1.00	1.32e-5
35	1.16e-3	9.57e-4	1.22	2.72e-07	2.72e-07	1.01	2.02e-6
40	1.16e-3	9.57e-4	1.22	8.22e-09	8.22e-09	1.00	2.31e-7
45	1.16e-03	9.57e-4	1.22	2.05e-09	2.05e-09	1.00	2.46e-08
50	1.16e-03	9.57e-4	1.22	1.93e-10	1.93e-10	1.00	1.94e-09

6.3 Eigenvalue Problems

Next, we consider the eigenvalue problem associated with the boundary value problem (6.6) of the Laplacian,

$$-\Delta u = \lambda u \quad \text{in } \Omega,\ u|_{\partial\Omega} = 0. \tag{6.18}$$

The corresponding variational formulation reads

$$a(u,\phi) = \lambda(u,\phi) \quad \forall \phi \in V = H_0^1(\Omega), \tag{6.19}$$

with the normalization $\|u\| = 1$. The corresponding Galerkin finite element approximation in $V_h \subset V$ reads

$$a(u_h,\phi_h) = \lambda_h(u_h,\phi_h) \quad \forall \phi_h \in V_h \tag{6.20}$$

with the normalization $\|u_h\| = 1$. The corresponding residual is given by

$$\begin{aligned}\rho(u_h,\lambda_h)(\psi) &:= \lambda_h(u_h,\psi) - a(u_h,\psi)\\ &= \sum_{K\in\mathbb{T}_h} \big\{ \big(R(u_h,\lambda_h),\psi\big)_K + \big(r(u_h),\psi\big)_{\partial K\setminus\partial\Omega} \big\},\end{aligned}$$

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

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

Theorem 6.2 *Let $\{\tilde{u}_h,\tilde{\lambda}_h\} \in V_h \times \mathbb{R}$, $\|\tilde{u}_h\| = 1$, be any approximation to the discrete eigenpair $\{u_h,\lambda_h\} \in V_h \times \mathbb{R}$ on the current mesh $\mathbb{T}_h$. Then, there holds*

$$(\tilde{\lambda}_h - \lambda)(1-\sigma_h) = \rho(\tilde{u}_h,\tilde{\lambda}_h)(u-\phi_h) + \rho(\tilde{u}_h,\tilde{\lambda}_h)(\phi_h) \tag{6.21}$$

for an arbitrary $\phi_h \in V_h$. *Here* $\sigma_h := \frac{1}{2}\|\tilde{u}_h - u\|^2$.

Proof [21] Observing $\|\tilde{u}_h\| = \|u\| = 1$, there holds

$$\begin{aligned}
&\rho(\tilde{u}_h, \tilde{\lambda}_h)(u - \phi_h) + \rho(\tilde{u}_h, \tilde{\lambda}_h)(\phi_h) \\
&\quad = \tilde{\lambda}_h(\tilde{u}_h, u) - a(\tilde{u}_h, u) \\
&\quad = (\tilde{\lambda}_h - \lambda)(\tilde{u}_h, u) + \lambda(\tilde{u}_h, u) - a(\tilde{u}_h, u) \\
&\quad = (\tilde{\lambda}_h - \lambda)(\tilde{u}_h, u) \\
&\quad = (\tilde{\lambda}_h - \lambda)\big(\tfrac{1}{2}\|\tilde{u}_h\|^2 + \tfrac{1}{2}\|u\|^2 - \tfrac{1}{2}\|\tilde{u}_h - u\|^2\big) \\
&\quad = (\tilde{\lambda}_h - \lambda)(1 - \sigma_h).
\end{aligned}$$

□

Remark 6.2 It is worth noting that:

1. The error representation has to be evaluated for a convergent sequence of approximate eigenfunctions: $\|\tilde{u}_h - u\|^2 \to 0$.
2. The evaluation of the error representation requires higher-order approximations $\hat{u}_h \approx u$ and $\hat{\sigma}_h \approx \sigma_h$ obtained, for example, from $\tilde{u}_h$ by post-processing as described above:

$$\tilde{\lambda}_h - \lambda \approx \frac{1}{1 - \hat{\sigma}_h}\big\{\rho(\tilde{u}_h, \tilde{\lambda}_h)(\hat{u}_h - \tilde{u}_h) + \rho(\tilde{u}_h, \tilde{\lambda}_h)(\tilde{u}_h)\big\}. \tag{6.22}$$

3. The second term on the right-hand side represents the deviation from Galerkin orthogonality and can be evaluated without any approximation.
4. The error representation (6.21) has a natural extension to non-symmetric eigenvalue problems (non-deficient eigenvalues):

$$\begin{aligned}
(\tilde{\lambda}_h - \lambda) \approx \frac{1}{1 - \hat{\sigma}_h}\big\{&\tfrac{1}{2}\rho(\tilde{u}_h, \tilde{\lambda}_h)\big(\hat{u}_h^* - \tilde{u}_h^*\big) + \tfrac{1}{2}\rho^*\big(\tilde{u}_h^*, \tilde{\lambda}_h\big)(\hat{u}_h - \tilde{u}_h) \\
&+ \tfrac{1}{2}\rho(\tilde{u}_h, \tilde{\lambda}_h)\big(\tilde{u}_h^*\big) + \tfrac{1}{2}\rho^*\big(\tilde{u}_h^*, \tilde{\lambda}_h\big)(\tilde{u}_h)\big\},
\end{aligned} \tag{6.23}$$

where $\hat{\sigma}_h := \frac{1}{2}(\tilde{u}_h - \hat{u}_h, \tilde{u}_h^* - \hat{u}_h^*)$, and $\tilde{u}_h^*$ is an approximation to the adjoint eigenfunction u^* corresponding to the eigenvalue λ. In the non-degenerate case, we can use the normalization $(u_h, u_h^*) = 1$ (see [15, 21]).

The results of various test computations reported in [21] demonstrate that our general approach to balancing discretization and iteration error also works well for symmetric as well as nonsymmetric eigenvalue problems. Based on the a posteriori error representations (6.21) or (6.23), we obtain effective stopping criteria for Krylov-space methods such as, for example, the Arnoldi method (see [22, 24]), which result in significant work savings.

6.4 Saddle Point Problems

The approach for simultaneous estimation of discretization and iteration errors introduced above can also be used for indefinite (linear) systems such as saddle point problems. We illustrate this for two different kinds of saddle point problems, the Stokes equation for modeling incompressible creeping viscous flow and the Karush-Kuhn-Tucker (KKT) system occurring as a first-order optimality condition of linear-quadratic optimal control problems.

6.4.1 Stokes Equations

The Stokes equation of fluid mechanics describes the behavior of a creeping incompressible fluid occupying a domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$,

$$\begin{aligned} &-\nu\Delta v + \nabla p = 0, \qquad \nabla \cdot v = 0 \quad \text{in } \Omega, \\ &v = 0 \quad \text{on } \Gamma_{\text{rigid}}, \qquad v = v^{\text{in}} \quad \text{on } \Gamma_{\text{in}}, \qquad \nu\partial_n v - pn = 0 \quad \text{on } \Gamma_{\text{out}}. \end{aligned} \tag{6.24}$$

The boundary is split like $\partial\Omega = \Gamma_{\text{rigid}} \cup \Gamma_{\text{in}} \cup \Gamma_{\text{out}}$, where Γ_{rigid} is the rigid part, Γ_{in} the inflow part, and Γ_{out} the usually artificial outflow part. For the meaning and properties of the Neumann-type outflow boundary condition (so-called "do-nothing" condition), we refer to [16]. Here, we consider the two-dimensional benchmark problem "channel flow around an obstacle" introduced in [23] (see Fig. 6.6). The quantity of interest is the drag coefficient

$$J(u) := \frac{2}{\bar{U}^2 D} \int_S n^T \big(2\nu\tau(v) - pI\big)e_1 \, \mathrm{ds},$$

where $u = \{v, p\}$, $\tau(v) := \frac{1}{2}(\nabla v + \nabla v^T)$ the strain tensor, n the outer normal unit vector along S, D the diameter of the obstacle, $\bar{U}$ the maximum inflow velocity, and e_1 the unit vector in the main flow direction. The variational formulation of the

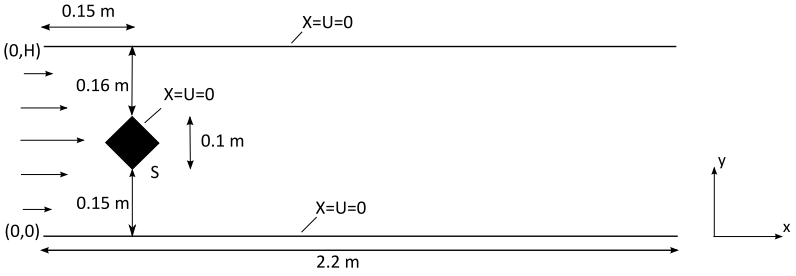

Fig. 6.6 Configuration of the flow example

Table 6.5 Iteration with *MG I* (iteration towards a round-off error level)

N	# Iter	$J(e)$	$\eta_h + \eta_{\text{it}}$	η_h	η_{it}	$I_{\text{eff}}^{\text{tot}}$
708	12	5.69e-05	9.19e-05	9.19e-05	2.03e-18	0.62
1 754	9	3.12e-05	2.81e-05	2.81e-05	1.05e-16	1.11
4 898	9	1.83e-05	1.21e-05	1.21e-05	2.20e-15	1.52
11 156	9	1.05e-05	7.01e-06	7.01e-06	9.49e-15	1.49
22 526	10	5.34e-06	3.77e-06	3.77e-06	8.36e-17	1.41
44 874	10	2.75e-06	2.12e-06	2.12e-06	3.39e-16	1.30
82 162	10	1.26e-06	1.09e-06	1.09e-06	4.29e-17	1.16
159 268	11	5.76e-07	6.11e-07	6.11e-07	1.26e-17	1.06
306 308	12	1.85e-07	2.98e-07	2.98e-07	8.74e-19	1.61

Table 6.6 Iteration with *MG II* (an adaptive stopping criterion)

N	# Iter	$J(e)$	$\eta_h + \eta_{\text{it}}$	η_h	η_{it}	$I_{\text{eff}}^{\text{tot}}$
708	2	5.69e-05	9.74e-05	9.17e-05	5.62e-06	0.59
1 754	2	3.12e-05	2.82e-05	2.81e-05	6.81e-08	1.11
4 898	2	1.83e-05	1.21e-05	1.21e-05	1.60e-08	1.52
11 156	2	1.05e-05	7.05e-06	7.01e-06	3.42e-08	1.49
22 526	2	5.34e-06	3.82e-06	3.77e-06	5.48e-08	1.39
44 874	2	2.75e-06	2.16e-06	2.12e-06	4.04e-08	1.28
82 162	2	1.27e-06	1.11e-06	1.09e-06	2.63e-08	1.14
159 268	2	5.76e-07	6.41e-07	6.10e-07	3.07e-08	0.90
306 308	2	1.86e-07	3.10e-07	2.97e-07	1.31e-08	0.60

problem (6.24) reads: Find $\{v, p\} \in (\hat{v}^{\text{in}} + H) \times L$ satisfying

$$\begin{aligned} \nu(\nabla v, \nabla \phi) - (p, \nabla \cdot \phi) &= (f, \phi) \quad \forall \phi \in H, \\ (\chi, \nabla \cdot v) &= 0 \qquad \forall \chi \in L, \end{aligned}$$

where $H := H_0^1(\Gamma_{\text{rigid}} \cup \Gamma_{\text{in}}; \Omega)^2$, $L := L^2(\Omega)$, and $\hat{v}^{\text{in}}$ is a suitable (solenoidal) extension of the boundary data.

The discretization uses equal-order (bilinear) Q_1 elements for velocity and pressure with additional pressure stabilization for circumventing the usual "inf-sup" stability condition,

$$\begin{aligned} \nu(\nabla v_h, \nabla \phi_h) - (p_h, \nabla \cdot \phi_h) &= (f, \phi_h) \quad \forall \phi_h \in H_h, \\ (\chi_h, \nabla \cdot v_h) + s_h(\chi_h, p_h) &= 0 \qquad \forall \chi_h \in L_h, \end{aligned} \tag{6.25}$$

where $V_h \subset V$ and $L_h \subset L$ are the finite element subspaces and $s_h(\chi_h, p_h)$ is a stabilizing form. For more details on pressure stabilization, we refer to the survey articles

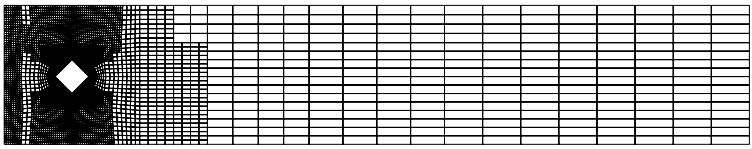

Fig. 6.7 A refined mesh with 4 898 knots in the flow example

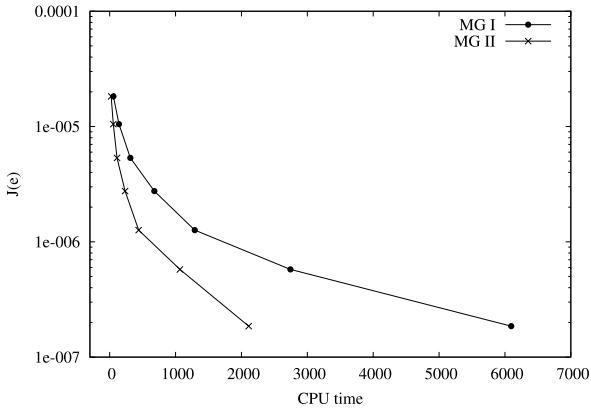

Fig. 6.8 Comparison of the CPU time used by the two *MG* variants *MG I* and *MG II*

[18, 19]. The discrete saddle point problem (6.25) is solved by an MG method using the canonical mesh transfer operations and "block ILU" smoothing (with $4+4$ smoothing steps). The computational results are shown in Tables 6.5 and 6.6 as well as Figs. 6.7 and 6.8.

6.4.2 The KKT System of Linear-Quadratic Optimization Problems

We consider the linear-quadratic optimization problem

$$\begin{aligned} J(u,q) &:= \tfrac{1}{2}\|u-\bar{u}\|^2 + \tfrac{1}{2}\alpha\|q\|^2 \longrightarrow \min, \\ -\Delta u &= f+q \quad \text{in } \Omega, \qquad u=0 \quad \text{on } \partial\Omega, \end{aligned} \tag{6.26}$$

on $\Omega := (0,1)^2 \subset \mathbb{R}^2$ with the force term f, prescribed target distribution $\bar{u}$, and distributed control q. The regularization parameter is taken as $\alpha = 10^{-3}$. This problem is solved by the Euler-Lagrange approach, which uses the Lagrangian functional

$$\mathscr{L}(u,q,\lambda) := J(u,q) + (f+q,\lambda) - (\nabla u, \nabla\lambda),$$

with the adjoint variable $\lambda \in V := H_0^1(\Omega)$. Then, for any optimal solution $\{u,q\} \in V \times Q := H_0^1(\Omega) \times L^2(\Omega)$ there exists an adjoint solution $\lambda \in V$ such that the

triplet $\{u, q, \lambda\} \in V \times Q \times V$ is a stationary point of the Lagrangian, i.e., it solves the following (linear) saddle point system:

$$\begin{aligned} (\nabla\phi, \nabla\lambda) - (u, \phi) &= -(\bar{u}, \phi) && \forall \phi \in V, \\ (\chi, \lambda) + \alpha(\chi, q) &= 0 && \forall \chi \in Q, \\ (\nabla u, \nabla\psi) - (q, \psi) &= (f, \psi) && \forall \psi \in V. \end{aligned} \tag{6.27}$$

This first-order necessary optimality condition is the so-called Karush-Kuhn-Tucker (KKT) system of the optimization problem.

For solving the KKT system (6.27), we use conforming bilinear Q_1 elements for all three variables $\{u, q, \lambda\}$. Denoting the corresponding finite element subspaces by $V_h \subset V$ and $Q_h \subset Q$, we obtain the discrete saddle point problem

$$\begin{aligned} (\nabla\phi_h, \nabla\lambda_h) - (u_h, \phi_h) &= -(\bar{u}, \phi_h) && \forall \phi_h \in V_h, \\ (\chi_h, \lambda_h) + \alpha(\chi_h, q_h) &= 0 && \forall \chi_h \in Q_h, \\ (\nabla u_h, \nabla\psi_h) - (q_h, \psi_h) &= (f, \psi_h) && \forall \psi_h \in V_h. \end{aligned} \tag{6.28}$$

This reads in a strong form as

$$\begin{aligned} -\Delta\lambda - u &= -\bar{u}, && \text{in } \Omega,\ \lambda|_{\partial\Omega} = 0, \\ \lambda + \alpha q &= 0, && \text{in } \Omega, \\ -\Delta u - q &= f, && \text{in } \Omega,\ u|_{\partial\Omega} = 0. \end{aligned} \tag{6.29}$$

This linear algebraic saddle point problem is again solved by a MG method using a block ILU iteration as a smoother.

Theorem 6.3 *Let $\{u, q, \lambda\} \in V \times Q \times V$ be the solution of the KKT system and $\{\tilde{u}_h, \tilde{q}_h, \tilde{\lambda}_h\} \in V_h \times Q_h \times V_h$ the approximative finite element solution of the discrete KKT system on the current mesh $\mathbb{T}_h$. Then, we have the error representation*

$$\begin{aligned} J(u, q) - J(\tilde{u}_h, \tilde{q}_h) = \tfrac{1}{2}\rho^*(\tilde{u}_h, \tilde{\lambda}_h)(u - \tilde{u}_h) + \tfrac{1}{2}\rho^q(\tilde{q}_h, \tilde{\lambda}_h)(q - \tilde{q}_h) \\ + \tfrac{1}{2}\rho(\tilde{u}_h, \tilde{q}_h)(\lambda - \tilde{\lambda}_h) + \rho(\tilde{u}_h, \tilde{q}_h)(\tilde{\lambda}_h), \end{aligned} \tag{6.30}$$

with the residuals

$$\begin{aligned} \rho^*(\tilde{u}_h, \tilde{\lambda}_h)(\phi) &:= (\tilde{u}_h - \bar{u}, \phi) - (\nabla\phi, \nabla\tilde{\lambda}_h), \\ \rho^q(\tilde{q}_h, \tilde{\lambda}_h)(\phi) &:= \alpha(\phi, \tilde{q}_h) + (\phi, \tilde{\lambda}_h), \\ \rho(\tilde{u}_h, \tilde{q}_h)(\phi) &:= (f + \tilde{q}_h, \phi) - (\nabla\tilde{u}_h, \nabla\phi). \end{aligned}$$

Proof For the proof, we refer to [17]. □

Remark 6.3 The choice of the cost functional $J(\cdot, \cdot)$ for error control may not be considered appropriate in the present case of a tracking problem where the particular

Table 6.7 *MG II* with block ILU smoothing, $\alpha = 10^{-3}$

N	E_{tot}	# Iter	E_h	η_h	I^h_{eff}	E_{it}	η_{it}	$I^{\text{it}}_{\text{eff}}$
25	9.35e-4	2	9.35e-4	1.83e-3	0.51	1.14e-07	1.97e-07	0.58
81	1.64e-4	2	1.78e-4	2.19e-4	0.82	1.42e-05	1.68e-05	0.85
289	3.75e-5	2	4.16e-5	4.39e-5	0.95	4.13e-06	4.33e-06	0.96
1 089	1.05e-5	2	1.02e-5	1.03e-5	0.99	3.48e-07	3.52e-07	0.99
3 985	2.67e-6	2	2.54e-6	2.55e-6	1.00	1.28e-07	1.28e-07	1.00
13 321	6.65e-7	2	6.48e-7	6.49e-7	1.00	1.63e-08	1.63e-08	1.00
47 201	1.76e-7	2	1.70e-7	1.69e-7	1.01	6.76e-09	6.77e-09	1.00
163 361	4.89e-8	2	4.69e-8	4.68e-8	1.01	1.97e-09	1.97e-09	1.00
627 697	1.23e-8	2	1.21e-8	1.21e-8	1.01	2.13e-10	2.13e-10	1.00

least-squares form of the functional is somewhat arbitrary. Instead, one may want to measure the solution accuracy rather in terms of some more relevant quantity depending on control and state, such as for example the norm $\|q - \tilde{q}_h\|_Q$ of the error in the control. This can be accomplished by utilizing an additional "outer" dual problem such as described in [5, 9].

We consider the example with the target distribution

$$\bar{u} = \tfrac{2\pi^2-1}{2\pi^2} \sin(\pi x) \sin(\pi y)$$

and the exact solution

$$u = -\frac{1}{2\pi^2} \sin(\pi x) \sin(\pi y), \qquad q = \frac{1}{2\alpha\pi^2} \sin(\pi x) \sin(\pi y),$$

$$\lambda = -\frac{1}{2\pi^2} \sin(\pi x) \sin(\pi y).$$

The forcing term f is accordingly adjusted. For simplicity, the discrete state and control spaces are chosen the same, $V_h = Q_h$, using isoparametric bilinear shape functions. For this test, we use the MG II algorithm with the stopping criterion

$$\eta_{\text{it}} \le \frac{1}{10}\, \eta_h.$$

First, we solve the discretized KKT system by the adaptive multigrid method using the V-cycle and again $4 + 4$-block-ILU smoothing steps on each level. Then, we use the multigrid method with only one undamped block-Jacobi smoothing step. The results are shown in Tables 6.7 and 6.8, where we use the abbreviations

$$E_{\text{tot}} := \big|J(u,q) - J(\tilde{u}_h, \tilde{q}_h)\big|, \qquad E_h := \big|J(u,q) - J(u_h, q_h)\big|,$$
$$E_{\text{it}} := \big|J(u_h,q_h) - J(\tilde{u}_h, \tilde{q}_h)\big|.$$

Table 6.8 *MG II* with block Jacobi smoothing, $\alpha = 10^{-3}$

N	E_{tot}	# Iter	E_h	η_h	I^h_{eff}	E_{it}	η_{it}	$I^{\text{it}}_{\text{eff}}$
25	9.44e-4	4	1.83e-3	9.35e-4	1.96	1.55e-5	8.99e-6	1.73
81	1.84e-4	5	2.20e-4	1.78e-4	1.23	7.59e-6	6.44e-6	1.18
289	4.36e-5	5	4.40e-5	4.16e-5	1.05	2.04e-6	1.96e-6	1.04
1 089	1.10e-5	4	1.03e-5	1.02e-5	1.01	8.53e-7	8.44e-7	1.01
3 985	2.69e-6	4	2.55e-6	2.56e-6	0.99	1.31e-7	1.30e-7	1.00
13 321	6.94e-7	4	6.47e-7	6.69e-7	0.96	2.51e-8	2.51e-8	1.00
47 201	1.95e-7	4	1.69e-7	1.90e-7	0.88	4.39e-9	4.40e-9	1.00
171 969	7.24e-8	3	4.42e-8	6.93e-8	0.63	3.07e-9	3.10e-9	0.99

We observe again a significant work saving by using the adaptive stopping criterion of the iteration.

6.5 The Nonlinear Case

Finally, we describe how our approach to the simultaneous estimation of the discretization and iteration errors extends to nonlinear variational problems of the form

$$A(u)(\psi) = F(\psi) \quad \forall \psi \in V, \; J(u) = ? \tag{6.31}$$

with a semi-linear "energy form" $A(\cdot)(\cdot)$ and a nonlinear output functional $J(\cdot)$ defined on the solution space V (both assumed to be sufficiently often differentiable). The starting point is the observation that any solution of the "primal" problem (6.31) corresponds to a stationary point of the Lagrangian functional $\mathscr{L}(u, z) := J(u) + F(z) - A(u)(z)$ with the dual variable $z \in V$ (Lagrangian multiplier). This results in the system

$$\begin{aligned} A(u)(\psi) &= F(\psi) && \forall \psi \in V, \\ A'(u)(\phi, z) &= J'(u)(\phi) && \forall \phi \in V. \end{aligned} \tag{6.32}$$

The finite element discretization of this system in spaces $V_h \subset V$ seeks primal and dual approximation $\{u_h, z_h\} \in V_h \times V_h$ satisfying

$$\begin{aligned} A(u_h)(\psi_h) &= F(\psi_h) && \forall \psi_h \in V_h, \\ A'(u_h)(\phi_h, z_h) &= J'(u_h)(\phi_h) && \forall \phi_h \in V_h. \end{aligned} \tag{6.33}$$

The corresponding primal and dual residuals are defined by

$$\rho(u_h)(\cdot) := F(\cdot) - A(u_h)(\cdot), \qquad \rho^*(u_h, z_h)(\cdot) := J'(u_h)(\cdot) - A'(u_h)(\cdot, z_h).$$

Theorem 6.4 *Let $\tilde{u}_h, \tilde{z}_h \in V_h$ be any approximations to the primal and dual discrete solutions $u_h, z_h \in V_h$ on the current mesh $\mathbb{T}_h$. Then, there holds*

$$J(u) - J(\tilde{u}_h) = \tfrac{1}{2}\rho(\tilde{u}_h)(z - \tilde{z}_h) + \tfrac{1}{2}\rho^*(\tilde{u}_h, \tilde{z}_h)(u - \tilde{u}_h) + \rho(\tilde{u}_h)(\tilde{z}_h) + \tilde{R}_h^{(3)} \quad (6.34)$$

with a remainder $\tilde{R}_h^{(3)}$ cubic in the errors $u - \tilde{u}_h$ and $z - \tilde{z}_h$.

Proof [20] For pairs $x = \{u, z\}$, we set $L(x) := \mathscr{L}(u, z)$. Then, with the abbreviation $\tilde{e}^z := u - \tilde{u}_h$, $\tilde{e}^z := z - \tilde{z}_h$, and $\tilde{e} := \{\tilde{e}^u, \tilde{e}^z\}$, there holds

$$\begin{aligned} J(u) - J(\tilde{u}_h) &= L(x) \underbrace{-F(z) + A(u)(z)}_{=0} - L(\tilde{x}_h) + \underbrace{F(\tilde{z}_h) - A(\tilde{u}_h)(\tilde{z}_h)}_{\neq 0} \\ &= \int_0^1 L'(\tilde{x}_h + s\tilde{e})(\tilde{e})\, ds + F(\tilde{z}_h) - A(\tilde{u}_h)(\tilde{z}_h). \end{aligned}$$

For the integral, we use the trapezoidal rule with integral remainder as follows:

$$\begin{aligned} J(u) - J(\tilde{u}_h) &= \tfrac{1}{2}\big\{\underbrace{L'(x)(\tilde{e})}_{=0} + \mathscr{L}'(\tilde{x}_h)(\tilde{e})\big\} \\ &\quad + \underbrace{\tfrac{1}{2}\int_0^1 L'''(\tilde{x}_h + s\tilde{e})(\tilde{e}, \tilde{e}, \tilde{e}) s(s-1)\, ds}_{=: \tilde{R}_h^{(3)}} + \underbrace{F(\tilde{z}_h) - A(\tilde{u}_h)(\tilde{z}_h)}_{= \rho(\tilde{u}_h)(\tilde{z}_h)} \\ &= \tfrac{1}{2}L'(\tilde{x}_h)(\tilde{e}) + \tilde{R}_h^{(3)} + \rho(\tilde{u}_h)(\tilde{z}_h) \\ &= \tfrac{1}{2}\big\{F(\tilde{e}^z) - A(\tilde{u}_h)(\tilde{e}^z) + J'(\tilde{u}_h)(\tilde{e}^u) - A'(\tilde{u}_h)(\tilde{e}^u, \tilde{z}_h)\big\} \\ &\quad + \tilde{R}_h^{(3)} + \rho(\tilde{u}_h)(\tilde{z}_h) \\ &= \tfrac{1}{2}\rho(\tilde{u}_h)(z - \tilde{z}_h) + \tfrac{1}{2}\rho^*(\tilde{u}_h, \tilde{z}_h)(u - \tilde{u}_h) + \tilde{R}_h^{(3)} + \rho(\tilde{u}_h)(\tilde{z}_h). \quad \square \end{aligned}$$

Remark 6.4 We make the following remarks:

1. The cubic remainder term $\tilde{R}_h^{(3)}$ is neglected or monitored by replacing $u - u_h^k \approx u_h^{k+1} - u_h^k$ and $z - z_h^k \approx z_h^{k+1} - z_h^k$.
2. For non-unique solutions the following a priori assumption $\{u_h, z_h\} \to \{u, z\}$ for $h \to 0$ is needed.
3. We have to solve the *linear* discrete dual problem

$$A'(u_h)(\phi_h, z_h) = J'(u_h)(\phi_h) \quad \forall \phi_h \in V_h. \quad (6.35)$$

4. The weights in the error representation are again approximated by patch-wise higher-order interpolation: $(z - \tilde{z}_h)_{|K} \approx (\tilde{I}_{2h}^{(2)}\tilde{z}_h - \tilde{z}_h)_{|K}$. The steps 1–4 are the essence of the **D**ual **W**eighted **R**esidual (DWR) method applied to the Galerkin finite element approximation of nonlinear problems.

5. The error representation (6.34) can be used to control the accuracy in the Newton iteration or in any other simple fixed point iteration for solving the algebraic problem (6.32).
6. If the approximative discrete solution $\tilde{u}_h$ is obtained by the Newton method, also an adaptive stopping criterion is needed for the inner linear solver of the single Newton steps. Such a strategy for simultaneous control of a discretization error, an outer *nonlinear* iteration error, and an inner *linear* iteration error can be developed on the basis of an a posteriori error representation by exploiting the structure of the Newton methods. For details, we refer to the forthcoming paper [20].

6.5.1 Numerical Example

We consider the following simple test problem: Compute $J(u) := u_1(a)$ for the solution $u \in V := H_0^1(\Omega)^2$ of the nonlinear system

$$\begin{aligned} -\Delta u_1 + 2u_2^2 &= 1, & u_1|_{\partial\Omega} &= 0, \\ -\Delta u_2 + u_1 u_2 &= 0, & u_2|_{\partial\Omega} &= 0. \end{aligned} \tag{6.36}$$

The configuration is shown in Fig. 6.9.

In this case the corresponding variational formulation reads

$$\begin{aligned} A(u)(\phi) &:= (\nabla u_1, \nabla \phi_1) + 2\big(u_1^2, \phi_1\big) + (\nabla u_2, \nabla \phi_2) + (u_1 u_2, \phi_2) \\ &= F(\phi) := (f, \phi) \quad \forall \phi \in V. \end{aligned} \tag{6.37}$$

For the discretization of the problem (6.37), we use again a standard finite element method with continuous Q_1 elements. The resulting nonlinear algebraic problems are solved by a damped Newton method with damping a factor $\theta = 0.5$,

$$A'\big(u_h^t\big)\big(u_h^{t+1}, \phi_h\big) = A'\big(u_h^t\big)\big(u_h^t, \phi_h\big) - \theta\big\{F(\phi_h) - A\big(u_h^t\big)(\phi_h)\big\}, \quad \forall \phi_h \in V_h. \tag{6.38}$$

We consider the following two different stopping criteria:

- *Newton I:* Reduction of initial Newton residual by factor 10^{-11};
- *Newton II:* Iteration error $\approx 10^{-1} \times$ discretization error.

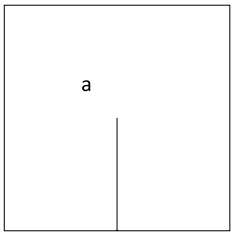

Fig. 6.9 Configuration of the nonlinear test problem: slit domain and point value evaluation

Table 6.9 *Newton I:* Iteration towards a "round-off error level" 10^{-11}

N	# Iter	$J(e)$	$\eta_h + \eta_{\text{it}}$	η_h	η_{it}	$I_{\text{eff}}^{\text{tot}}$
85	31	3.31e-03	1.49e-03	1.49e-03	1.69e-11	2.22
297	29	1.24e-03	5.78e-04	5.78e-04	7.14e-11	2.13
897	29	5.46e-04	2.30e-04	2.30e-04	7.26e-11	2.38
2063	29	2.43e-04	9.59e-05	9.59e-05	7.32e-11	2.56
4537	27	1.14e-04	4.34e-05	4.34e-05	2.94e-10	2.63
9969	27	5.28e-05	2.15e-05	2.15e-05	2.94e-10	2.44
21389	27	2.23e-05	1.03e-05	1.03e-05	2.94e-10	2.17
39549	27	7.58e-06	5.36e-06	5.36e-06	2.94e-10	1.41

Table 6.10 *Newton II:* An adaptive stopping criterion

N	# Iter	$J(e)$	$\eta_h + \eta_{\text{it}}$	η_h	η_{it}	$I_{\text{eff}}^{\text{tot}}$
85	8	3.31e-03	1.63e-03	1.49e-03	1.41e-04	2.13
297	10	1.24e-03	6.15e-04	5.77e-04	3.74e-05	2.08
897	11	5.46e-04	2.49e-04	2.30e-04	1.90e-05	2.27
2063	13	2.43e-04	1.01e-04	9.59e-05	4.79e-06	2.44
4537	14	1.14e-04	4.58e-05	4.34e-05	2.40e-06	2.56
9969	15	5.28e-05	2.27e-05	2.15e-05	1.20e-06	2.38
21389	16	2.23e-05	1.09e-05	1.03e-05	6.03e-07	2.08
39549	17	7.58e-06	5.66e-06	5.36e-06	3.01e-07	1.39

The linear subproblems are solved by an MG iteration with

- Smoother: Jacobi with damping factor 0.5;
- Stopping criterion: Reduction of the initial multigrid residual by factor 10^{-11}.

The obtained results are shown in Tables 6.9 and 6.10. Again, we observe significant work savings through the adaptive stopping criterion. The effectivity indices are relatively close to one, even on coarser meshes, which demonstrates the sharpness of our error indicators. However, we observe slight underestimation on all meshes.

6.6 Conclusion and Outlook

Goal-oriented adaptivity by the DWR method is in principle possible for all problems formulated within a variational setting. Though largely of heuristic nature the DWR method provides a general guideline for treating even most complex nonlinear systems. However, its theoretical justification in any particular case requires additional assumptions and hard work. In this way the discretization error and the algebraic iteration error, linear as well as nonlinear, can be simultaneously controlled

leading to effective stopping criteria and significant work savings. Current developments into the same direction are a posteriori control of the following additional "variational crimes":

- Quadrature error,
- Boundary approximation,
- Stabilization error ("inf-sup" and "transport" stabilization),
- Domain approximation (truncation of unbounded domains),
- Various modeling errors.

This will be the subject of forthcoming papers.

References

1. Babuška I, Strouboulis T (2001) The finite element method and its reliability. Clarendon Press, New York
2. Bangerth W, Rannacher R (2003) Adaptive finite element methods for differential equations. Birkhäuser, Basel
3. Bank RE, Dupont T (1981) An optimal order process for solving finite element equations. Math Comput 36(153):35–51
4. Becker R (1998) An adaptive finite element method for the Stokes equations including control of the iteration error. In: Enumath 97: 2nd European conference on numerical mathematics and advanced applications, Heidelberg, 1997. World Scientific, River Edge, pp 609–620
5. Becker R, Braack M, Meidner D, Rannacher R, Vexler B (2007) Adaptive finite element methods for pde-constrained optimal control problems. In: Jäger W, Rannacher R, Warnatz J (eds) Reactive flow, diffusion and transport. Springer, Berlin, pp 177–205
6. Becker R, Johnson C, Rannacher R (1995) Adaptive error control for multigrid finite element methods. Computing 55(4):271–288
7. Becker R, Rannacher R (1996) A feed-back approach to error control in finite element methods: basic analysis and examples. East-West J Numer Math 4(4):237–264
8. Becker R, Rannacher R (2001) An optimal control approach to a posteriori error estimation in finite element methods. Acta Numer 10:1–102
9. Becker R, Vexler B (2004) A posteriori error estimation for finite element discretization of parameter identification problems. Numer Math 96(3):435–459
10. Bramble JH (1993) Multigrid methods. Pitman research notes in mathematics, vol 294. Longman, Harlow
11. Bramble JH, Pasciak JE (1993) New estimates for multilevel algorithms including the V-cycle. Math Comput 60(202):447–471
12. Bramble JH, Pasciak JE, Wang JP, Xu J (1991) Convergence estimates for multigrid algorithms without regularity assumptions. Math Comput 57(195):23–45
13. Ciarlet PG (2002) The finite element method for elliptic problems. Classics appl math, vol 40. SIAM, Philadelphia
14. Hackbusch W (1985) Multigrid methods and applications. Springer, Berlin
15. Heuveline V, Rannacher R (2001) A posteriori error control for finite element approximations of elliptic eigenvalue problems. Adv Comput Math 15(1–4):107–138
16. Heywood J, Rannacher R, Turek S (1996) Artificial boundaries and flux and pressure conditions for the incompressible Navier-Stokes equations. Int J Numer Methods Fluids 22(5):325–352
17. Meidner D, Rannacher R, Vihharev J (2009) Goal-oriented error control of the iterative solution of finite element equations. J Numer Math 17(2):143–172

18. Rannacher R (2000) Finite element methods for the incompressible Navier-Stokes equations. In: Galdi GP, Heywood JG, Rannacher R (eds) Fundamental directions in mathematical fluid mechanics. Birkhäuser, Basel, pp 191–293
19. Rannacher R (2004) Incompressible viscous flow. In: Stein E, de Borst R, Hughes TJR (eds) Encyclopedia of computational mechanics. Vol. 3. Fluids. Wiley, Chichester
20. Rannacher R, Vihharev J (2011) Adaptive finite element analysis of nonlinear problems: balancing of discretization and iteration error. Preprint, University of Heidelberg
21. Rannacher R, Westenberger A, Wollner W (2010) Adaptive finite element solution of eigenvalue problems: balancing of discretization and iteration error. J Numer Math 18(4):303–327
22. Saad Y (1980) Variations on Arnoldi's method for computing eigenelements of large unsymmetric matrices. Linear Algebra Appl 34:269–295
23. Schäfer M, Turek S (1996) Benchmark computations of laminar flow around a cylinder. In: Hirschel EH (ed) Flow simulation with high-performance computers II. NNFM, vol 52. Vieweg, Braunschweig, pp 547–566
24. Sorensen DC (2002) Numerical methods for large eigenvalue problems. Acta Numer 11:519–584
25. Verfürth R (1996) A review of a posteriori error estimation and adaptive Mesh-refinement techniques. Wiley-Teubner, Chichester