MATHEMATICS OF COMPUTATION
VOLUME 62, NUMBER 206
APRIL 1994, PAGES 445–475

A POSTERIORI ERROR ESTIMATES FOR NONLINEAR PROBLEMS. FINITE ELEMENT DISCRETIZATIONS OF ELLIPTIC EQUATIONS

R. VERFÜRTH

ABSTRACT. We give a general framework for deriving a posteriori error estimates for approximate solutions of nonlinear problems. In a first step it is proven that the error of the approximate solution can be bounded from above and from below by an appropriate norm of its residual. In a second step this norm of the residual is bounded from above and from below by a similar norm of a suitable finite-dimensional approximation of the residual. This quantity can easily be evaluated, and for many practical applications sharp explicit upper and lower bounds are readily obtained. The general results are then applied to finite element discretizations of scalar quasi-linear elliptic partial differential equations of 2nd order, the eigenvalue problem for scalar linear elliptic operators of 2nd order, and the stationary incompressible Navier-Stokes equations. They immediately yield a posteriori error estimates, which can easily be computed from the given data of the problem and the computed numerical solution and which give global upper and local lower bounds on the error of the numerical solution.

1. INTRODUCTION

The efficiency of a numerical method for the solution of partial differential equations strongly depends on the choice of an "optimal" discretization, the use of a fast and efficient algorithm for the solution of the discrete problem, and a simple, but reliable method for judging the quality of the numerical solution obtained. These three objectives are often interdependent. The first and last one are related to the problem of a posteriori error estimation, i.e., of extracting from the given data of the problem and the computed numerical solution reliable bounds on the error of the numerical solution. Of course, the computation of the a posteriori error estimates should be much less costly than the solution of the original discrete problem.

Within the framework of finite element methods various strategies of a posteriori error estimation have been devised during the last 15–20 years (cf., e.g., [2, 3, 20, 27] and the literature cited there). They can roughly be classified as follows:

(1) *residual estimates*: Estimate the error of the computed numerical solution by a suitable norm of its residual with respect to the strong form of the differential equation (cf., e.g., [4, 5, 9, 19, 21, 25, 27]).

(2) *solution of local problems*: Solve locally discrete problems similar to,

Received by the editor December 14, 1992.
1991 *Mathematics Subject Classification*. Primary 65N30, 65N15, 65J15.

©1994 American Mathematical Society
0025-5718/94 $1.00 + $.25 per page

but simpler than, the original problem and use appropriate norms of the local solutions for error estimation (cf., e.g., [7, 8, 18, 22, 25, 27]).

(3) *sharp a priori error estimates*: Derive sharp a priori error estimates and use suitable higher-order difference quotients of the computed numerical solution to estimate the higher-order derivatives appearing in the a priori error estimates (cf., e.g., [15, 16]).

(4) *averaging methods*: Use some local averaging technique for error estimation (cf., e.g., [6, 21, 29, 30]).

For a certain class of problems and discretizations it was proven in [28] that the methods (1) and (2) are equivalent in the sense that, up to multiplicative constants, they yield the same upper and lower bounds on the error of the numerical solution (cf. also [6, 13, 21] for the comparison of different error estimators). In this context it should be noted that, in order to be efficient, an a posteriori error estimation should yield upper and lower bounds on the error. Clearly, upper bounds are sufficient to ensure that the numerical solution achieves a prescribed tolerance. Lower bounds, however, are essential to guarantee that the error is not overestimated and that its local distribution is correctly resolved. Often, only upper bounds are established in the literature.

Various methods are used for constructing a posteriori error estimators and for proving that they yield upper and/or lower bounds on the error. These methods often depend on a particular class of problems and discretizations. A close inspection, however, reveals that they have certain principles in common. It is the aim of this paper to give a rather general framework that allows one to construct a posteriori error estimators and to prove that they yield upper and lower bounds on the error. In this general context we are satisfied with proving that the upper and lower bounds differ by a multiplicative constant which is independent of the mesh size. We neither intend to derive optimal estimates for this constant nor to prove efficiency of the error estimators, i.e., that the ratio of the true and the estimated error asymptotically tends to 1. This latter question is addressed for linear problems in e.g. [2, 3, 4, 5, 6, 7, 13, 14].

We consider in §§2–4 nonlinear equations of the form

$$F(u) = 0 \tag{1.1}$$

and corresponding discretizations of the form

$$F_h(\overset{\circ}{u}_h) = 0. \tag{1.2}$$

Here, $F \in C^1(X, Y^*)$ and $F_h \in C(X_h, Y_h^*)$, $X_h \subset X$ and $Y_h \subset Y$ are finite-dimensional subspaces of the Banach spaces X and Y, and $*$ denotes the dual of a Banach space.

If $u_0 \in X$ is a solution of equation (1.1) such that $DF(u_0)$ is an isomorphism of X onto Y^* and DF is Lipschitz continuous at u_0, we prove in Proposition 2.1 that

$$\underline{c}\|F(u)\|_{Y^*} \le \|u - u_0\|_X \le \bar{c}\|F(u)\|_{Y^*} \tag{1.3}$$

holds for all u in a suitable neighborhood of u_0. The constants $\underline{c}$ and $\bar{c}$ depend on $DF(u_0)$ and $DF(u_0)^{-1}$. The proof of Proposition 2.1 is straightforward. The conditions on F can be weakened considerably (cf. Remark 2.3). Inequality (1.3) is a local result. That means that it can be applied to solutions of equation (1.2) only if they are sufficiently close to u_0, i.e., if the discretization is "sufficiently fine". This is not surprising since we are dealing with general nonlinear problems, which may have a large variety of solutions. If problem (1.1) is linear, i.e., DF is constant, inequality (1.3) of course holds for all $u \in X$.

In §3 we briefly outline how the results of §2 can be extended to branches of solutions of equation (1.1), including singular points such as simple limit and bifurcation points. The generalization to the case of a regular branch of solutions, i.e., situations covered by the implicit function theorem, is straightforward. The case of a simple limit or bifurcation point can be reduced as in [12] to the case of a regular branch of solutions by suitably blowing up the spaces X and Y and modifying the function F. For practical applications it is important that the additional spaces are finite-dimensional. Thus, the cost for evaluating the residual of the modified function is essentially determined by the cost for evaluating the residual of F.

In §4, we estimate the residual $\|F(u_h)\|_{Y^*}$, where u_h is an approximate solution of equation (1.2). To this end, we introduce a restriction operator $R_h\colon Y \to Y_h$, a finite-dimensional subspace $\widetilde{Y}_h \subset Y$, and an approximation $\widetilde{F}_h\colon X_h \to Y^*$ of F at u_h which are coupled via inequality (4.1). For practical applications, the construction of R_h and $\widetilde{F}_h$ is rather straightforward. Usually, $\widetilde{F}_h(u_h)$ is obtained by locally projecting $F(u_h)$ onto suitable finite-dimensional spaces. This corresponds to the well-known technique of locally freezing the coefficients of a differential operator. The choice of $\widetilde{Y}_h$ on the other hand is less obvious. It is, however, considerably simplified by the auxiliary results of §5 (see also below). We then prove in Proposition 4.1 that, up to multiplicative constants and additive correction terms, $\|F(u_h)\|_{Y^*}$ is bounded from below and from above by $\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}$. The latter can be evaluated quite easily since its computation is equivalent to a finite-dimensional maximization problem. Moreover, sharp explicit bounds on $\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}$ are readily obtained for many practical applications. When applying the general results to finite element methods, the aforementioned multiplicative constants essentially depend on the element geometry and on the polynomial degree of the finite element functions. In principle, they can be estimated explicitly. The aforementioned correction terms consist of the following quantities:

(1) the residual $\|F_h(u_h)\|_{Y_h^*}$ of the discrete problem (1.2),

(2) the consistency error $\|F(u_h) - F_h(u_h)\|_{Y_h^*}$ of the discretization, and

(3) a term which measures the quality of the approximation of $F(u_h)$ by $\widetilde{F}_h(u_h)$.

The first quantity can easily be estimated from u_h and the given data. The second one can be bounded a priori. For many practical applications one can finally prove that the third quantity is a higher-order perturbation when compared with $\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}$.

In this section we also give a framework which covers some of the a posteriori error estimators based on the solution of auxiliary local problems, such as the one described in [4, 5], and which shows that these estimators are equivalent to the residual a posteriori error estimator considered before.

As already mentioned, we establish in §5 some auxiliary results which simplify the construction of $\widetilde{Y}_h$. The main result is of the form (cf. Lemma 5.1)

$$(1.4) \qquad 0 \leq \alpha \leq \inf_{u \in V_S \setminus \{0\}} \sup_{v \in V_S \setminus \{0\}} \frac{\int_S u \psi_S v}{\|u\|_{L^p(S)} \|v\|_{L^q(S)}} \leq 1.$$

Here, $1 \leq p \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, S is either a simplex in $\mathbb{R}^n$ or a face of such a simplex, V_S is a finite-dimensional space of functions defined on S, and ψ_S is a cutoff function. It is important to note that the constant α is independent of S. Lemma 5.1 is a generalization of Lemma 4.1 in [28]. Thanks to inequality (1.4), one can show that for finite element methods, $\widetilde{Y}_h$ can be chosen as the space of all linear combinations of functions $\psi_S v$, where $v \in V_S$ and S varies through all elements and their faces.

In §§6–8 we apply the general results of the previous sections to finite element approximations of scalar quasi-linear elliptic partial differential equations of 2nd order, the eigenvalue problem for scalar linear elliptic differential operators of 2nd order, and the stationary incompressible Navier-Stokes equations (cf. Propositions 6.1, 6.3, 6.4, 6.5, 7.1, 8.1, and 8.4). In all examples we obtain upper and lower bounds for the finite element error in terms of a residual a posteriori error estimator. This error estimator essentially consists of the elementwise error of the finite element functions with respect to the strong form of the differential equation and of jumps across inter-element boundaries of that boundary operator which naturally links the strong and weak forms of the differential equation. Some of the results of §§6–8 are completely new, others are generalizations of, and improvements upon, results previously obtained in [4, 5, 7, 8, 9, 19, 25, 27, 28].

2. Error estimates for isolated solutions

Let X, Y be two Banach spaces with norms $\|\cdot\|_X$ and $\|\cdot\|_Y$. For any element $u \in X$ and any real number $R > 0$ set $B(u, R) := \{v \in X : \|u-v\|_X < R\}$. We denote by $\mathscr{L}(X, Y)$ and $\operatorname{Isom}(X, Y) \subset \mathscr{L}(X, Y)$ the Banach space of continuous linear maps of X in Y equipped with the operator norm $\|\cdot\|_{\mathscr{L}(X,Y)}$, and the open subset of linear homeomorphisms of X onto Y. By $Y^* := \mathscr{L}(Y, \mathbb{R})$ and $\langle \cdot , \cdot \rangle$ we denote the dual space of Y and the corresponding duality pairing. Finally, $A^* \in \mathscr{L}(Y^*, Y^*)$ denotes the adjoint of a given linear operator $A \in \mathscr{L}(Y, Y)$.

Let $F \in C^1(X, Y^*)$ be a given continuously differentiable function. The following proposition yields a posteriori error estimates for elements in a neighborhood of a solution of equation (1.1).

Proposition 2.1. *Let $u_0 \in X$ be a regular solution of equation* (1.1); *i.e., $DF(u_0) \in \operatorname{Isom}(X, Y^*)$. Assume that DF is Lipschitz continuous at u_0; i.e., there is an $R_0 > 0$ such that*

$$\gamma := \sup_{u \in B(u_0, R_0)} \frac{\|DF(u) - DF(u_0)\|_{\mathscr{L}(X,Y^*)}}{\|u - u_0\|_X} < \infty.$$

Set

$$R := \min\{R_0, \gamma^{-1} \|DF(u_0)^{-1}\|^{-1}_{\mathscr{L}(Y^*,X)}, 2\gamma^{-1} \|DF(u_0)\|_{\mathscr{L}(X,Y^*)}\}.$$

Then the following error estimates hold for all $u \in B(u_0, R)$:

$$\begin{aligned} \frac{1}{2}\|DF(u_0)\|^{-1}_{\mathscr{L}(X,Y^*)}\|F(u)\|_{Y^*} \\ \le \|u-u_0\|_X \le 2\|DF(u_0)^{-1}\|_{\mathscr{L}(Y^*,X)}\|F(u)\|_{Y^*}. \end{aligned} \tag{2.1}$$

Proof. Let $u \in B(u_0, R)$. We then have

$$u - u_0 = DF(u_0)^{-1}\left\{F(u) + \int_0^1 [DF(u_0) - DF(u_0 + t(u-u_0))](u-u_0)\,dt\right\}$$

and thus

$$\begin{aligned} &\|u-u_0\|_X \\ &\quad\le \|DF(u_0)^{-1}\|_{\mathscr{L}(Y^*,X)} \\ &\qquad\cdot\left\{\|F(u)\|_{Y^*} + \int_0^1 \|DF(u_0) - DF(u_0+t(u-u_0))\|_{\mathscr{L}(X,Y^*)}\|u-u_0\|_X\,dt\right\} \\ &\quad\le \|DF(u_0)^{-1}\|_{\mathscr{L}(Y^*,X)}\left\{\|F(u)\|_{Y^*} + \frac{1}{2}\gamma\|u-u_0\|_X^2\right\} \\ &\quad\le \|DF(u_0)^{-1}\|_{\mathscr{L}(Y^*,X)}\|F(u)\|_{Y^*} + \frac{1}{2}\|u-u_0\|_X. \end{aligned}$$

This yields the second inequality in (2.1).

On the other hand, we have for all $\varphi \in Y$ with $\|\varphi\|_Y = 1$

$$\begin{aligned} \langle F(u), \varphi\rangle &= \langle DF(u_0)(u-u_0), \varphi\rangle \\ &\quad+ \left\langle \int_0^1 [DF(u_0+t(u-u_0)) - DF(u_0)](u-u_0)\,dt, \varphi\right\rangle \end{aligned} \tag{2.2}$$

and thus

$$\begin{aligned} \|F(u)\|_{Y^*} &\le \|DF(u_0)\|_{\mathscr{L}(X,Y^*)}\|u-u_0\|_X \\ &\quad+ \int_0^1 \|DF(u_0+t(u-u_0)) - DF(u_0)\|_{\mathscr{L}(X,Y^*)}\|u-u_0\|_X\,dt \\ &\le \|DF(u_0)\|_{\mathscr{L}(X,Y^*)}\|u-u_0\|_X + \frac{1}{2}\gamma\|u-u_0\|_X^2 \\ &\le 2\|DF(u_0)\|_{\mathscr{L}(X,Y^*)}\|u-u_0\|_X. \end{aligned}$$

This proves the first inequality in (2.1). □

Remark 2.2. In the examples of §§6–8, X and Y are closed subspaces of suitable Sobolev spaces of functions defined on an open set $\Omega \subset \mathbb{R}^n$. When considering in equation (2.2) only functions φ with support in a given open subset $\omega \subset \Omega$, one then often obtains lower bounds for $u - u_0$ restricted to ω. □

Remark 2.3. The conditions about F can be weakened. Assume, e.g., that $F \in C(X, Y^*)$, $F(u_0) = 0$, and that there are an $R > 0$ and two monotonically increasing homeomorphisms ϱ, σ of $[0, \infty)$ onto itself such that

$$\varrho(\|u-u_0\|_X) \le \|F(u)\|_{Y^*} \le \sigma(\|u-u_0\|_X) \quad \forall u \in B(u_0, R). \tag{2.3}$$

We then trivially have

$$\sigma^{-1}(\|F(u)\|_{Y^*}) \le \|u-u_0\|_X \le \varrho^{-1}(\|F(u)\|_{Y^*}) \quad \forall u \in B(u_0, R).$$

The first inequality in (2.3) is satisfied if, e.g., F is strongly monotone in a neighborhood of u_0. The second inequality in (2.3) holds if, e.g., F is Hölder continuous at u_0. □

3. Error estimates for branches of solutions

In this section we briefly outline how the results of the previous section may be extended to branches of solutions of equation (1.1), including simple limit and bifurcation points. To this end, we assume that $X = \mathbb{R}^m \times V$, $m \ge 1$, and that $u_0 = (\lambda_0, v_0)$ is a solution of equation (1.1).

We first consider the case that u_0 is a *regular point*, i.e.,

$$D_V F(u_0) \in \operatorname{Isom}(V, Y^*).$$

The implicit function theorem then implies that there are neighborhoods I of λ_0 in $\mathbb{R}^m$ and U of v_0 in V and a continuous map $\lambda \to v_\lambda$ from I into U such that $v_{\lambda_0} = v_0$ and every $u_\lambda := (\lambda, v_\lambda)$ is a solution of equation (1.1) with $D_V F(u_\lambda) \in \operatorname{Isom}(V, Y^*)$. Assume that there is an $R_0^* > 0$ such that

$$\gamma^* := \sup_{\lambda \in I} \sup_{v \in B(v_\lambda, R_0^*)} \frac{\|D_V F(\lambda, v) - D_V F(\lambda, v_\lambda)\|_{\mathscr{L}(V, Y^*)}}{\|v - v_\lambda\|_V} < \infty,$$

and set

$$R^* := \min\{R_0^*, \gamma^{*^{-1}} \sup_{\lambda \in I} \|D_V F(u_\lambda)^{-1}\|^{-1}_{\mathscr{L}(Y^*, V)}, 2\gamma^{*^{-1}} \sup_{\lambda \in I} \|D_V F(u_\lambda)\|_{\mathscr{L}(V, Y^*)}\}.$$

With the same arguments as in the proof of Proposition 2.1 we then obtain for all $\lambda \in I$ and all $v \in B(v_\lambda, R^*)$ the estimates

$$\begin{aligned} &\frac{1}{2}\|D_V F(u_\lambda)\|^{-1}_{\mathscr{L}(V, Y^*)} \|F(\lambda, v)\|_{Y^*} \\ &\quad \le \|v - v_\lambda\|_V \le 2\|D_V F(u_\lambda)^{-1}\|_{\mathscr{L}(Y^*, V)} \|F(\lambda, v)\|_{Y^*}. \end{aligned} \tag{3.1}$$

As described in [12], the case where u_0 is not a regular point, but a simple limit or bifurcation point, may be reduced to the case of a regular point by suitably blowing up the spaces X and Y and modifying the function F. For completeness, we briefly describe this procedure.

Consider first the case that u_0 is a *simple limit point*; i.e., $DF(u_0)$ is a Fredholm operator of X onto Y^* with index m and $\operatorname{Range}(DF(u_0)) = Y^*$ but $D_V F(u_0) \notin \operatorname{Isom}(V, Y^*)$. Choose a linear operator $B \in \mathscr{L}(X, \mathbb{R}^m)$ with $\ker(B) \cap \ker(DF(u_0)) = \{0\}$ and define $\Phi \in C^1(\mathbb{R}^m \times X, \mathbb{R}^m \times Y^*)$ by

$$\Phi(t, u) := (B(u - u_0) - t, F(u)).$$

Then, $(0, u_0)$ is a regular point of Φ (with respect to the parameter t), and we are back to the situation described in the first part of this section. Since B is linear, conditions about the Lipschitz continuity of $D\Phi$ reduce to those on DF. Equation (3.1) yields in this case estimates of the form

$$\begin{aligned} &\underline{c}\{\|B(u-u_0) - t\|_{\mathbb{R}^m} + \|F(u)\|_{Y^*}\} \le \|\lambda - \lambda_t\|_{\mathbb{R}^m} + \|v - v_t\|_V \\ &\qquad \le \bar{c}\{\|B(u-u_0) - t\|_{\mathbb{R}^m} + \|F(u)\|_{Y^*}\} \end{aligned} \tag{3.2}$$

for all t in a suitable neighborhood of 0 and all $u = (\lambda, v)$ in a suitable neighborhood of $u_t = (\lambda_t, v_t)$. Here, $t \to u_t$ is a regular branch of solutions of $\Phi(t, u) = 0$. Note, that Bu_0 is often known explicitly and that the estimation of $\|B(u-u_0)-t\|_{\mathbb{R}^m}$ is straightforward, since it is a low-dimensional maximization problem. The term $\|F(u)\|_{Y^*}$, on the other hand, may be estimated by the methods of the next section, as in the case of regular solutions.

Next, we consider the case of a *simple bifurcation from the trivial branch.* That is, we assume that $u_0 = (\lambda_0, 0)$ and that $D_V F(u_0)$ is a Fredholm operator with index 0 and $\dim\ker(D_V F(u_0)) = 1$. Choose a $w_0 \in \ker(D_V F(u_0))\backslash\{0\}$ and a linear functional $l \in \mathscr{L}(V, \mathbb{R})$ with $l(w_0) = 1$. Define the function $\Phi \in C(\mathbb{R} \times X, \mathbb{R} \times Y^*)$ by

$$\Phi(t, u) := \begin{cases} (l(v) - 1, \frac{1}{t}F(\lambda, tv)), & t \neq 0, \ u = (\lambda, v) \in X, \\ (l(v) - 1, D_V F(\lambda, 0)v), & t = 0, \ u = (\lambda, v) \in X. \end{cases}$$

Conditions about the Lipschitz continuity of $D\Phi$ now reduce to those on D^2F. Obviously, we have $\Phi(0, \tilde{u}_0) = 0$, where $\tilde{u}_0 := (\lambda_0, w_0)$. If F is of class C^2 in a neighborhood of u_0 and $D^2_{\lambda v}F(u_0)w_0 \notin \operatorname{Range} D_V F(u_0)$, we conclude that $\tilde{u}_0$ is a regular point, and we are once more back to the situation described in the first part of this section. Equation (3.1) now yields estimates of the form

$$\begin{aligned} \underline{c}\{|l(w) - 1| + \|D_V F(\lambda, 0)w\|_{Y^*}\} &\leq \|\lambda - \lambda_0\|_{\mathbb{R}^m} + \|w - w_0\|_V \\ &\leq \bar{c}\{|l(w) - 1| + \|D_V F(\lambda, 0)w\|_{Y^*}\} \end{aligned} \tag{3.3}$$

for all (λ, w) in a suitable neighborhood of $\tilde{u}_0$ and

$$\begin{aligned} \underline{c}\left\{|l(w) - 1| + \left\|\frac{1}{t}F(\lambda, tw)\right\|_{Y^*}\right\} &\leq \|\lambda - \lambda_t\|_{\mathbb{R}^m} + \|w - w_t\|_V \\ &\leq \bar{c}\left\{|l(w) - 1| + \left\|\frac{1}{t}F(\lambda, tw)\right\|_{Y^*}\right\} \end{aligned} \tag{3.4}$$

for all $t \neq 0$ in a neighborhood of 0 and all (λ, w) in a suitable neighborhood of $\tilde{u}_t = (\lambda_t, w_t)$. Here, $t \to \tilde{u}_t$ is a regular branch of solutions of $\Phi(t, u) = 0$. Note that the constants in equations (3.3), (3.4) now depend on second derivatives of F.

Finally, we consider the case of a *simple bifurcation point*; i.e., $D_V F(u_0)$ is a Fredholm operator of index 0 and $q := \dim(\ker DF(u_0)) - m \geq 1$. Choose a basis $\varphi_1^*, \dots, \varphi_q^*$ of $Y^*\backslash \operatorname{Range}(DF(u_0))$, set $\widehat{X} := \mathbb{R}^q \times X$, $\hat{u}_0 := (0, u_0)$, and define the function $\widehat{F} \in C^1(\widehat{X}, Y)$ by

$$\widehat{F}(\hat{u}) := F(u) - \sum_{i=1}^{q} f_i\varphi_i^* \quad \forall \hat{u} = (f, u) \in \widehat{X}.$$

Obviously, we have $\widehat{F}(\hat{u}_0) = 0$. Moreover, $D\widehat{F}(\hat{u}_0)$ is a Fredholm operator with index $m + q$ and $\operatorname{Range}(D\widehat{F}(\hat{u}_0)) = Y^*$. Replacing X, u_0, and F by $\widehat{X}$, $\hat{u}_0$, and $\widehat{F}$, respectively, we are thus back to the situation considered in the second part of this section.

4. Estimation of the residual

Let $X_h \subset X$ and $Y_h \subset Y$ be finite-dimensional subspaces and $F_h \in C(X_h, Y_h^*)$ be an approximation of F. We want to estimate $\|F(u_h)\|_{Y^*}$, where $u_h \in X_h$ is an approximate solution of equation (1.2).

In what follows, $c, c_0, c_1, \ldots$ denote various constants which are independent of h.

Proposition 4.1. *Let $u_h \in X_h$ be an approximate solution of equation* (1.2); *i.e., $\|F_h(u_h)\|_{Y_h^*}$ is "small". Assume that there are a restriction operator $R_h \in \mathscr{L}(Y, Y_h)$, a finite-dimensional subspace $\widetilde{Y}_h \subset Y$, and an approximation $\widetilde{F}_h : X_h \to Y^*$ of F at u_h such that*

$$\|(\mathrm{Id}_Y - R_h)^* \widetilde{F}_h(u_h)\|_{Y^*} \le c_0 \|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}. \tag{4.1}$$

Then the following estimates hold:

$$\begin{aligned} \|F(u_h)\|_{Y^*} \le c_0 \|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*} &+ \|(\mathrm{Id}_Y - R_h)^*[F(u_h) - \widetilde{F}_h(u_h)]\|_{Y^*} \\ &+ \|R_h\|_{\mathscr{L}(Y, Y_h)} \|F(u_h) - F_h(u_h)\|_{Y_h^*} \\ &+ \|R_h\|_{\mathscr{L}(Y, Y_h)} \|F_h(u_h)\|_{Y_h^*} \end{aligned} \tag{4.2}$$

and

$$\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*} \le \|F(u_h)\|_{\widetilde{Y}_h^*} + \|F(u_h) - \widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}. \tag{4.3}$$

Remark 4.2. In the examples of §§6–8, X_h and Y_h are suitable finite element spaces. The choice of R_h is then quite natural. $\widetilde{F}_h(u_h)$ is obtained by projecting $F(u_h)$ elementwise onto suitable finite-dimensional spaces. This construction is also rather standard. The main difficulty is to find a space $\widetilde{Y}_h$ such that inequality (4.1) is satisfied. This task is simplified by the auxiliary results of §5. The second terms on the right-hand sides of equations (4.2) and (4.3) measure the quality of the approximation $\widetilde{F}_h(u_h)$ to $F(u_h)$. Usually, they are higher-order terms when compared with $\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}$. The term $\|F(u_h) - F_h(u_h)\|_{Y_h^*}$ is the consistency error of the discretization. The term $\|F_h(u_h)\|_{Y_h^*}$ measures the residual of the algebraic equation (1.2) and can easily be evaluated.

Proof of Proposition 4.1. Consider an arbitrary element $\varphi \in Y$ with $\|\varphi\|_Y = 1$. We then have

$$\begin{aligned} \langle F(u_h), \varphi\rangle &= \langle \widetilde{F}_h(u_h), \varphi - R_h\varphi\rangle + \langle F(u_h) - \widetilde{F}_h(u_h), \varphi - R_h\varphi\rangle \\ &\quad + \langle F(u_h) - F_h(u_h), R_h\varphi\rangle + \langle F_h(u_h), R_h\varphi\rangle \\ &\le \|(\mathrm{Id}_Y - R_h)^* \widetilde{F}_h(u_h)\|_{Y^*} + \|(\mathrm{Id}_Y - R_h)^*[F(u_h) - \widetilde{F}_h(u_h)]\|_{Y^*} \\ &\quad + \|R_h\|_{\mathscr{L}(Y, Y_h)} \|F(u_h) - F_h(u_h)\|_{Y_h^*} + \|R_h\|_{\mathscr{L}(Y, Y_h)} \|F_h(u_h)\|_{Y_h^*}. \end{aligned}$$

Together with inequality (4.1), this proves estimate (4.2). Estimate (4.3) follows from the triangle inequality. □

When combining Propositions 2.1 and 4.1 we obtain a residual a posteriori error estimator. The following proposition together with Proposition 2.1 yields a framework for some of those a posteriori error estimators which are based on the solution of auxiliary local problems, such as the one described in [4, 5].

Proposition 4.3. *Let $u_h \in X_h$ be an approximate solution of equation* (1.2). *Assume that there are finite-dimensional subspaces $\widehat{X}_h \subset X$ and $\widehat{Y}_h \subset Y$ and a linear operator $B \in \mathrm{Isom}(\widehat{X}_h, \widehat{Y}_h^*)$ such that $\widetilde{Y}_h \subset \widehat{Y}_h$ and*

$$\|\widetilde{F}_h(u_h)\|_{\widehat{Y}_h^*} \le c_1 \|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}. \tag{4.4}$$

Let $\hat{u}_h \in \widehat{X}_h$ *be the unique solution of*

$$(4.5)\qquad \langle B\hat{u}_h\,,\,\varphi\rangle = \langle \widetilde{F}_h(u_h)\,,\,\varphi\rangle \quad \forall\varphi \in \widehat{Y}_h.$$

Then the following estimates hold:

$$(4.6)\qquad \|B\|^{-1}_{\mathscr{L}(\widehat{X}_h\,,\,\widehat{Y}_h^*)}\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*} \le \|u_h\|_{\widehat{X}_h} \le c_1\|B^{-1}\|_{\mathscr{L}(\widehat{Y}_h^*\,,\,\widehat{X}_h)}\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}.$$

Proof. Since $B \in \operatorname{Isom}(\widehat{X}_h\,,\,\widehat{Y}_h^*)$, we immediately obtain from equation (4.5) the estimate

$$(4.7)\qquad \|B\|^{-1}_{\mathscr{L}(\widehat{X}_h\,,\,\widehat{Y}_h^*)}\|\widetilde{F}_h(u_h)\|_{\widehat{Y}_h^*} \le \|u_h\|_{\widehat{X}_h} \le \|B^{-1}\|_{\mathscr{L}(\widehat{Y}_h^*\,,\,\widehat{X}_h)}\|\widetilde{F}_h(u_h)\|_{\widehat{Y}_h^*}.$$

Together with inequality (4.4), this proves the upper bound of inequality (4.6). Since $\widetilde{Y}_h \subset \widehat{Y}_h$, we have

$$\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*} \le \|\widetilde{F}_h(u_h)\|_{\widehat{Y}_h^*}.$$

Together with inequality (4.7), this proves the lower bound of inequality (4.6). □

Remark 4.4. Usually, B is some approximation of $DF(u_h)$. The construction of $\widehat{Y}_h$ and the proof of inequality (4.4) are similar to the construction of $\widetilde{Y}_h$ and the proof of inequality (4.1) and are simplified by the auxiliary results of §5. Once B and $\widehat{Y}_h$ are chosen, the construction of $\widehat{X}_h$ is quite obvious from the condition that $B \in \operatorname{Isom}(\widehat{X}_h\,,\,\widehat{Y}_h^*)$.

5. Auxiliary results

Let Ω be a bounded, connected, open domain in $\mathbb{R}^n$, $n \ge 2$, with polyhedral boundary Γ. For any open subset $\omega \subset \Omega$ with Lipschitz boundary γ, we denote by $W^{k,s}(\omega)$, $k \in \mathbb{N}$, $1 \le s \le \infty$, $L^s(\omega) := W^{0,s}(\omega)$, and $L^s(\gamma)$ the usual Sobolev and Lebesgue spaces equipped with the standard norms $\|\cdot\|_{k,s;\omega} := \|\cdot\|_{W^{k,s}(\omega)}$ and $\|\cdot\|_{s;\gamma} := \|\cdot\|_{L^s(\gamma)}$ (cf. [1]). If $\omega = \Omega$, we omit the index ω. We use the same notation for the corresponding norms of vector-valued functions.

Let $\mathscr{T}_h$, $h > 0$, be a family of partitions of Ω into n-simplices, which satisfies the following conditions:

(1) Any two simplices in $\mathscr{T}_h$ are either disjoint or share a complete smooth submanifold of their boundaries.

(2) The ratio h_T/ϱ_T is bounded from above independently of $T \in \mathscr{T}_h$ and $h > 0$.

Here, h_T, ϱ_T, and h_E denote the diameter of $T \in \mathscr{T}_h$, the diameter of the largest ball inscribed into T, and the diameter of a face E of T. Note, that condition (2) allows the use of locally refined meshes and that it implies that the ratio h_T/h_E, for all $T \in \mathscr{T}_h$ and all faces E of T, is bounded from above and from below by constants which are independent of h, T, and E.

Denote by $\mathscr{E}_h$ the set of all faces of all $T \in \mathscr{T}_h$. The set $\mathscr{E}_h$ may be decomposed as $\mathscr{E}_h = \mathscr{E}_{h,\Omega} \cup \mathscr{E}_{h,\Gamma}$, $\mathscr{E}_{h,\Omega} \cap \mathscr{E}_{h,\Gamma} = \varnothing$, where $\mathscr{E}_{h,\Gamma}$ denotes the set of all faces lying on Γ. Given an $E \in \mathscr{E}_h$, we denote by ω_E the union of all simplices in $\mathscr{T}_h$ having E as a face. Similarly, ω_T, $T \in \mathscr{T}_h$, is the union of all simplices sharing a face with T. For any $E \in \mathscr{E}_h$ and any piecewise continuous function φ, we denote by $[\varphi]_E$ the jump of φ across E in a fixed direction. Here, φ is continued by 0 outside Ω and the direction is given by the exterior normal of Γ if $E \in \mathcal{E}_{h,\Gamma}$.

For $k \in \mathbb{N}$, we define

$$S_h^{k,-1} := \{\varphi : \Omega \to \mathbb{R} : \varphi\,|_T \in \Pi_k \ \forall T \in \mathcal{T}_k\}, \qquad S_h^{k,0} := S_h^{k,-1} \cap C(\overline{\Omega}).$$

Here, Π_k, $k \geq 0$, is the space of polynomials of degree at most k. Moreover, we denote by $\pi_{k,S}$, $S \in \mathcal{T}_h \cup \mathcal{E}_h$, the L^2-projection of $L^1(S)$ onto $\Pi_k\,|_S$.

Using standard scaling arguments for finite elements, we finally conclude from [11] that there is an "interpolation" operator $I_h \colon L^1(\Omega) \to S_h^{1,0}$ which satisfies the following error estimates for all $T \in \mathcal{T}_h$, $E \in \mathcal{E}_h$, and $1 \leq q \leq \infty$:

$$(5.1)\qquad \|\varphi - I_h\varphi\|_{k,q;T} \leq c_1 h_T^{l-k} \|\varphi\|_{l,q;\widetilde{\omega}_T} \quad \forall 0 \leq k \leq l \leq 1,\ \varphi \in W^{l,q}(\widetilde{\omega}_T),$$

$$(5.2)\qquad \|\varphi - I_h\varphi\|_{q;E} \leq c_2 h_E^{1-1/q} \|\varphi\|_{1,q;\widetilde{\omega}_E} \quad \forall \varphi \in W^{1,q}(\widetilde{\omega}_E),$$

where $\widetilde{\omega}_T$ and $\widetilde{\omega}_E$ denote the union of all elements having a nonempty intersection with T and E, respectively. Here and in what follows, we adopt the usual convention that $1/\infty := 0$.

Denote by $\widehat{T} := \{\hat{x} \in \mathbb{R}^n : \sum_{i=1}^n \hat{x}_i \leq 1,\ \hat{x}_j \geq 0,\ 1 \leq j \leq n\}$ the reference n-simplex. Set $\widehat{E} := \widehat{T} \cap \{\hat{x} \in \mathbb{R}^n : \hat{x}_n = 0\}$, and let $\hat{x}_{\widehat{T}}$ and $\hat{x}_{\widehat{E}}$ be the barycenters of $\widehat{T}$ and $\widehat{E}$, respectively. The following conditions uniquely define two functions $\psi_{\widehat{T}}, \psi_{\widehat{E}} \in C^\infty(\widehat{T}, \mathbb{R})$:

$$\begin{aligned} &\psi_{\widehat{T}} \in \Pi_{n+1}, \quad \psi_{\widehat{T}}(\hat{x}_{\widehat{T}}) = 1, \quad \psi_{\widehat{T}} = 0 \ \text{ on } \partial\widehat{T}, \\ &\psi_{\widehat{E}} \in \Pi_n, \quad \psi_{\widehat{E}}(\hat{x}_{\widehat{E}}) = 1, \quad \psi_{\widehat{E}} = 0 \ \text{ on } \partial\widehat{T}\backslash\widehat{E}. \end{aligned}$$

Note, that the above conditions, in particular, imply that

$$0 \leq \psi_{\widehat{T}} \leq 1, \quad 0 \leq \psi_{\widehat{E}} \leq 1 \qquad \text{in } \widehat{T}.$$

We define a continuation operator $\widehat{P} \colon L^\infty(\widehat{E}) \to L^\infty(\widehat{T})$ by

$$\widehat{P}\hat{u}(\hat{x}_1, \ldots, \hat{x}_n) := \hat{u}(\hat{x}_1, \ldots, \hat{x}_{n-1}) \quad \forall \hat{x} \in \widehat{T},\ \hat{u} \in L^\infty(\widehat{E}).$$

Finally, $V_{\widehat{T}} \subset L^\infty(\widehat{T})$ and $V_{\widehat{E}} \subset L^\infty(\widehat{E})$ are two arbitrary finite-dimensional spaces, which are kept fixed throughout this section.

Let $T \in \mathcal{T}_h$ be an arbitrary n-simplex and $E \subset \partial T$ be a face of T. There is an invertible affine mapping $F_T \colon \widehat{T} \to T$, $\hat{x} \to x := F_T(\hat{x}) = b_T + B_T\hat{x}$ such that $\widehat{T}$ is mapped onto T and $\widehat{E}$ is mapped onto E. Denote by B'_T the matrix which is obtained from B_T by discarding its last column, and set $\beta_T := \det(B'^t_T B'_T)^{1/2}$, the Gram determinant of the transformation $\widehat{E} \to E$. Set

$$\begin{aligned} \psi_T &:= \psi_{\widehat{T}} \circ F_T^{-1}, & \psi_E &:= \psi_{\widehat{E}} \circ F_T^{-1}, \\ V_T &:= \{\hat{u} \circ F_T^{-1} : \hat{u} \in V_{\widehat{T}}\}, & V_E &:= \{\hat{\sigma} \circ F_T^{-1} : \hat{\sigma} \in V_{\widehat{E}}\}. \end{aligned}$$

Finally, we define the continuation operator $P \colon L^\infty(E) \to L^\infty(T)$ by

$$P\sigma := [\widehat{P}\sigma \circ F_T] \circ F_T^{-1}.$$

In what follows, p, q are two fixed real numbers with $1 \leq p \leq \infty$ and $\frac{1}{p} + \frac{1}{q} = 1$, and $|||\cdot|||$ denotes the spectral norm on $\mathbb{R}^{n\times n}$.

Lemma 5.1. *There are constants* $c_1, \dots, c_7$, *which only depend on the spaces* $V_{\widehat{T}}$ *and* $V_{\widehat{E}}$, *the number* p, *and the ratio* h_T/ϱ_T, *such that the following inequalities hold for all* $u \in V_T$ *and all* $\sigma \in V_E$:

$$c_1\|u\|_{0,p;T} \le \sup_{v\in V_T} \frac{\int_T u\psi_T v}{\|v\|_{0,q;T}} \le \|u\|_{0,p;T}, \tag{5.3}$$

$$c_2\|\sigma\|_{p;E} \le \sup_{\tau\in V_E} \frac{\int_E \sigma\psi_E \tau}{\|\tau\|_{q;E}} \le \|\sigma\|_{p;E}, \tag{5.4}$$

$$c_3 h_T^{-1}\|\psi_T u\|_{0,q;T} \le \|\nabla(\psi_T u)\|_{0,q;T} \le c_4 h_T^{-1}\|\psi_T u\|_{0,q;T}, \tag{5.5}$$

$$c_5 h_T^{-1}\|\psi_E P\sigma\|_{0,q;T} \le \|\nabla(\psi_E P\sigma)\|_{0,q;T} \le c_6 h_T^{-1}\|\psi_E P\sigma\|_{0,q;T}, \tag{5.6}$$

$$\|\psi_E P\sigma\|_{0,q;T} \le c_7 h_T^{1/q}\|\sigma\|_{q;E}. \tag{5.7}$$

Proof. The upper bounds of equations (5.3), (5.4) immediately follow from Hölder's inequality and $0 \le \psi_T \le 1$, $0 \le \psi_E \le 1$.

In order to prove the lower bound of equation (5.3), one easily checks that the mapping

$$\hat{u} \to \sup_{\hat{v}\in V_{\widehat{T}}} \frac{\int_{\widehat{T}} \hat{u}\psi_{\widehat{T}}\hat{v}}{\|\hat{v}\|_{0,q;\widehat{T}}}$$

defines a norm on $V_{\widehat{T}}$. Since $\dim V_{\widehat{T}} < \infty$, there is a constant $\hat{c} > 0$ such that

$$\hat{c}\|\hat{u}\|_{0,p;\widehat{T}} \le \sup_{\hat{v}\in V_{\widehat{T}}} \frac{\int_{\widehat{T}} \hat{u}\psi_{\widehat{T}}\hat{v}}{\|\hat{v}\|_{0,q;\widehat{T}}} \quad \forall \hat{u} \in V_{\widehat{T}}.$$

Now, take an arbitrary $u \in V_T$. Set $\hat{u} := u \circ F_T \in V_{\widehat{T}}$ and choose a $\hat{w} \in V_{\widehat{T}}$ such that

$$\|\hat{w}\|_{0,q;\widehat{T}} = 1 \quad \text{and} \quad \int_{\widehat{T}} \hat{u}\psi_{\widehat{T}}\hat{w} \ge \hat{c}\|\hat{u}\|_{0,p;\widehat{T}}.$$

With $w := \hat{w} \circ F_T^{-1}$ we then obtain

$$\begin{aligned}\sup_{v\in V_T} \frac{\int_T u\psi_T v}{\|v\|_{0,q;T}} &\ge \frac{\int_T u\psi_T w}{\|w\|_{0,q;T}} = |\det B_T|^{1-1/q} \int_{\widehat{T}} \hat{u}\psi_{\widehat{T}}\hat{w} \\ &\ge \hat{c}|\det B_T|^{1/p}\|\hat{u}\|_{0,p;\widehat{T}} = \hat{c}\|u\|_{0,p;T}.\end{aligned}$$

The proof of the lower bound of equation (5.4) is completely analogous. One only has to replace $|\det B_T|$ by β_T.

The mappings

$$\hat{u} \to \|\nabla(\psi_{\widehat{T}}\hat{u})\|_{0,q;\widehat{T}} \quad \text{and} \quad \hat{\sigma} \to \|\nabla(\psi_{\widehat{E}}\widehat{P}\hat{\sigma})\|_{0,q;\widehat{T}}$$

define norms on $V_{\widehat{T}}$ and $V_{\widehat{E}}$. Since $\psi_{\widehat{T}}$ and $\psi_{\widehat{E}}$ vanish at the vertices of $\widehat{T}$, and since $\dim V_{\widehat{T}} < \infty$ and $\dim V_{\widehat{E}} < \infty$, these norms are equivalent to $\|\psi_{\widehat{T}}\hat{u}\|_{0,q;\widehat{T}}$ and $\|\psi_{\widehat{E}}\widehat{P}\hat{\sigma}\|_{0,q;\widehat{T}}$, respectively. Estimates (5.5), (5.6) now follow in the usual way by transforming to $\widehat{T}$, using the equivalence of norms there, and transforming back to T.

With the same arguments as above we finally conclude that there is a constant $\tilde{c} > 0$ such that

$$\|\psi_{\widehat{E}}\widehat{P}\hat{\sigma}\|_{0,q;\widehat{T}} \le \tilde{c}\|\hat{\sigma}\|_{q;\widehat{E}} \quad \forall \hat{\sigma} \in V_{\widehat{E}}.$$

Since

$$|\det B_T| \le |||B_T|||^n \le ch_T^n, \qquad \beta_T^{-1} \le |||B_T^{-1}|||^{n-1} \le ch_T^{1-n},$$

this implies for all $\sigma \in V_E$

$$\begin{aligned}\|\psi_E P\sigma\|_{0,q;T} &= |\det B_T|^{1/q} \|\psi_{\widehat{E}} \widehat{P}\hat{\sigma}\|_{0,q;\widehat{T}} \le \tilde{c}|\det B_T|^{1/q}\|\hat{\sigma}\|_{q;\widehat{E}} \\ &= \tilde{c}|\det B_T|^{1/q}\beta_T^{-1/q}\|\sigma\|_{q;E} \le c_7 h_T^{1/q}\|\sigma\|_{q;E}. \quad \square\end{aligned}$$

Remark 5.2. The estimates of Lemma 5.1 also hold for "slightly curved" simplices. More precisely, assume that the transformation F_T is no longer affine, but that it still is a diffeomorphism. Let $A_T\colon \widehat{T} \to \mathbb{R}^n$ be the invertible affine mapping which is uniquely determined by the condition that $A_T^{-1} \circ F_T$ leaves the vertices of $\widehat{T}$ invariant. Denote by α_T the Gram determinant of the transformation of $\widehat{E}$ induced by A_T. A simple perturbation argument then shows that the estimates of Lemma 5.1 remain valid, provided

$$\| \, |||I - DF_T^{-1}DA_T||| \, \|_{0,\infty;\widehat{T}}, \qquad \| \, |||I - DA_T^{-1}DF_T||| \, \|_{0,\infty;\widehat{T}},$$
$$\|1 - |\det DF_T|^{-1}|\det DA_T| \, \|_{0,\infty;\widehat{T}}, \qquad \|1 - \beta_T^{-1}\alpha_T\|_{0,\infty;\widehat{T}}$$

are smaller than a positive threshold which only depends on the constants in the corresponding estimates on $\widehat{T}$. □

Thanks to Lemma 5.1, we may construct in the next section spaces $\widetilde{Y}_h$ and $\widehat{Y}_h$ satisfying the conditions of Propositions 4.1 and 4.3 by considering all linear combinations of functions $\psi_T v$ and $\psi_E P\sigma$, where v and σ vary in suitable spaces V_T and V_E, respectively, and T and E run through all simplices and faces of the finite element partition.

Note that Lemma 5.1 does not depend on the fact that $\psi_{\widehat{T}}$ and $\psi_{\widehat{E}}$ are polynomials. This special choice has only been made for convenience.

6. Scalar quasi-linear elliptic equations of 2nd order

Consider the boundary value problem

$$\begin{aligned}-\nabla\cdot\underline{a}(x, u, \nabla u) &= b(x, u, \nabla u) \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \Gamma,\end{aligned} \tag{6.1}$$

where $b \in C(\Omega \times \mathbb{R} \times \mathbb{R}^n, \mathbb{R})$ and $\underline{a} \in C^1(\Omega \times \mathbb{R} \times \mathbb{R}^n, \mathbb{R}^n)$ are such that the matrix $A(x, y, z) := (\frac{1}{2}(\partial_{z_j} a_i(x, y, z) + \partial_{z_i} a_j(x, y, z))_{1\le i,j\le n}$ is positive definite for all $x \in \Omega$, $y \in \mathbb{R}$, $z \in \mathbb{R}^n$.

Under suitable growth conditions on $\underline{a}$, b, and their derivatives there are real numbers $1 < r, q < \infty$ such that the weak formulation of problem (6.1) fits into the framework of §2 with

$$X := \{u \in W^{1,r}(\Omega) \colon u = 0 \text{ on } \Gamma\}, \qquad \|\cdot\|_X := \|\cdot\|_{1,r},$$
$$Y := \{\varphi \in W^{1,q}(\Omega) : u = 0 \text{ on } \Gamma\}, \qquad \|\cdot\|_Y := \|\cdot\|_{1,q},$$
$$\langle F(u), \varphi\rangle := \int_\Omega \underline{a}(x, u, \nabla u)\nabla\varphi - \int_\Omega b(x, u, \nabla u)\varphi.$$

Denote by $p := \frac{q}{q-1}$ the dual exponent of q. Note that $DF(u) \in \operatorname{Isom}(X, Y^*)$ if the linear boundary value problem

$$\begin{aligned}&-\nabla\cdot(A(x, u, \nabla u)\nabla v) - \nabla\cdot(\partial_y \underline{a}(x, u, \nabla u)v) \\ &\qquad - \nabla_z b(x, u, \nabla u)\cdot\nabla v - \partial_y b(x, u, \nabla u)v = f \quad \text{in } \Omega, \\ &\qquad\qquad v = 0 \quad \text{on } \Gamma\end{aligned}$$

admits for each $f \in Y^*$ a unique weak solution $v \in X$ which depends continuously on f.

Some examples of problems falling into the present category are given by:

(1) The equations of prescribed mean curvature:

$$\begin{aligned} \underline{a}(x, u, \nabla u) &:= [1 + \|\nabla u\|^2]^{-1/2}\nabla u, \\ b(x, u, \nabla u) &:= f(x) \in L^2(\Omega), \\ r &:= q := 2. \end{aligned}$$

(2) The α-Laplacian:

$$\begin{aligned} \underline{a}(x, u, \nabla u) &:= \|\nabla u\|^{\alpha-2}\nabla u, \qquad \alpha > 1, \\ b(x, u, \nabla u) &:= f(x) \in L^p(\Omega), \\ r &:= q := \alpha. \end{aligned}$$

(3) The subsonic flow of an irrotational, ideal, compressible gas:

$$\begin{aligned} \underline{a}(x, u, \nabla u) &:= \left[1 - \frac{\gamma - 1}{2}\|\nabla u\|^2\right]^{1/(\gamma-1)} \nabla u, \qquad \gamma > 1, \\ b(x, u, \nabla u) &:= f(x) \in L^p(\Omega), \\ r &:= q := \frac{2\gamma}{\gamma - 1}. \end{aligned}$$

(4) The stationary heat equation with convection and nonlinear diffusion coefficient:

$$\begin{aligned} \underline{a}(x, u, \nabla u) &:= k(u)\nabla u, \\ b(x, u, \nabla u) &:= f - \mathbf{c} \cdot \nabla u, \end{aligned}$$

$$f \in L^\infty(\Omega), \quad \mathbf{c} \in C(\Omega, \mathbb{R}^n), \quad k \in C^2(\mathbb{R}),$$
$$k(s) \geq \alpha > 0, \quad |k^{(l)}(s)| \leq \gamma, \quad \forall s \in \mathbb{R}, \ l = 0, 1, 2,$$
$$r := p \in (2, 4).$$

(5) Bratu's equation:

$$\begin{aligned} \underline{a}(x, u, \nabla u) &:= \nabla u, \\ b(x, u, \nabla u) &:= \lambda e^u, \qquad \lambda > 0, \\ r &:= p > n. \end{aligned}$$

(6) A nonlinear eigenvalue problem:

$$\begin{aligned} \underline{a}(x, u, \nabla u) &:= \nabla u, \\ b(x, u, \nabla u) &:= \lambda u - u^\beta, \qquad \beta \geq n, \\ r &:= p \geq n. \end{aligned}$$

Example (2) fits into the framework of Proposition 2.1 if $\alpha \geq 2$. If $1 < \alpha < 2$, the corresponding function F is no longer differentiable. However, it still fits into the framework of Remark 2.3 with

$$\varrho(t) = \underline{c}\{\|u_0\|_X + t\}^{\alpha-2}t, \qquad \sigma(t) = \bar{c}t^{\alpha-1}$$

(cf. [10, §5.3]).

In example (5) there is a critical parameter $\lambda^* > 0$ such that the problem admits two weak solutions if $0 < \lambda < \lambda^*$, exactly one weak solution if $\lambda = \lambda^*$, and no solution if $\lambda > \lambda^*$. The solution corresponding to $\lambda = \lambda^*$ is a turning point and fits into the framework of the second part of §3 (cf. [12]).

Example (6) always admits the trivial solution. If λ is a simple eigenvalue of the Laplacian, there is a simple bifurcation which fits into the framework of the third part of §3 (cf. [12]).

We do not specify the discretization of problem (6.1) in detail. We only assume that $X_h \subset X \cap W^{1,\infty}(\Omega)$ and $Y_h \subset Y \cap W^{1,\infty}(\Omega)$ are finite element spaces corresponding to $\mathcal{T}_h$ consisting of affinely equivalent elements in the sense of [10], and that $S_h^{1,0} \cap Y \subset Y_h$.

In order to construct R_h, $\widetilde{F}_h$ and $\widetilde{Y}_h$, we define two integers k, l and approximations $\underline{a}_h$ of $\underline{a}$ and b_h of b as follows:

$$\underline{a}_h(x, u_h, \nabla u_h) := \begin{cases} \underline{a}(x, u_h, \nabla u_h), & \text{if } \underline{a}(x, v_h, \nabla v_h) \in S_h^{k,-1}\ \forall v_h \in X_h, \\ \sum\limits_{T\in\mathcal{T}_h} \pi_{1,T} a(x, u_h, \nabla u_h),\ k := 1, & \text{otherwise,} \end{cases}$$

$$b_h(x, u_h, \nabla u_h) := \begin{cases} b(x, u_h, \nabla u_h), & \text{if } b(x, v_h, \nabla v_h) \in S_h^{l,-1}\ \forall v_h \in X_h, \\ \sum\limits_{T\in\mathcal{T}_h} \pi_{0,T} b(x, u_h, \nabla u_h),\ l := 0 & \text{otherwise.} \end{cases}$$

Here, $u_h \in X_h$ is arbitrary. Now, $\widetilde{F}_h$ is defined in the same way as F with $\underline{a}$ and b replaced by $\underline{a}_h$ and b_h, respectively, $R_h := I_h$, and

$$\widetilde{Y}_h := \operatorname{span}\{\psi_T v,\ \psi_E P\sigma : v \in \Pi_m|_T,\ \sigma \in \Pi_k|_E,\ T \in \mathcal{T}_h,\ E \in \mathcal{E}_{h,\Omega}\},$$

where $m := \max\{k-1, l\}$.

Put, for abbreviation,

$$\begin{aligned} \varepsilon_T := \Big\{ h_T^p \| -\nabla\cdot(\underline{a}(\cdot, u_h, \nabla u_h) - \underline{a}_h(\cdot, u_h, \nabla u_h)) \\ - (b(\cdot, u_h, \nabla u_h) - b_h(\cdot, u_h, \nabla u_h))\|_{0,p;T}^p \\ + \sum_{E\subset\partial T\setminus\Gamma} h_E \|[n\cdot(\underline{a}(\cdot, u_h, \nabla u_h) \\ - \underline{a}_h(\cdot, u_h, \nabla u_h))]_E\|_{p;E}^p \Big\}^{1/p} \quad \forall T \in \mathcal{T}_h, \end{aligned} \tag{6.2}$$

$$\begin{aligned} \eta_T := \Big\{ h_T^p \| -\nabla\cdot\underline{a}_h(\cdot, u_h, \nabla u_h) - b_h(\cdot, u_h, \nabla u_h)\|_{0,p;T}^p \\ + \sum_{E\subset\partial T\setminus\Gamma} h_E \|[n\cdot\underline{a}_h(\cdot, u_h, \nabla u_h)]_E\|_{p;E}^p \Big\}^{1/p} \quad \forall T \in \mathcal{T}_h. \end{aligned} \tag{6.3}$$

The quantity ε_T obviously measures the quality of the approximation of $\underline{a}$ and b by $\underline{a}_h$ and b_h, respectively, and can be estimated explicitly. Below, we will show that $\|(\mathrm{Id}_Y - R_h)^*[F(u_h) - \widetilde{F}_h(u_h)]\|_{Y^*}$ and $\|F(u_h) - \widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*}$ are bounded from above by $c\{\sum_{T\in\mathcal{T}_h} \varepsilon_T^p\}^{1/p}$.

Note that

$$\varepsilon_T = h_T \|f - \pi_{0,T} f\|_{0,p;T} \quad \forall T \in \mathcal{T}_h,$$

if $X_h \subset S_h^{1,0}$ in examples (1)–(3), that

$$\varepsilon_T \le c h_T^{2(1-n/p)} \|\nabla u_h\|_{0,p;T} \quad \forall T \in \mathcal{T}_h$$

if $X_h \subset S_h^{1,0}$ in example (4), that

$$\varepsilon_T \le c h_T^2 \|\nabla u_h\|_{0,p;T} \exp(\|u_h\|_{0,\infty;T}) \quad \forall T \in \mathcal{T}_h$$

if X_h consists of piecewise polynomials in example (5), and that

$$\varepsilon_T = 0 \quad \forall T \in \mathcal{T}_h$$

if X_h consists of piecewise polynomials and $\beta \in \mathbb{N}$ in example (6).

Using integration by parts elementwise, we obtain for all $\varphi \in Y$

$$\begin{aligned}\langle F(u_h), \varphi\rangle = &\sum_{T \in \mathcal{T}_h} \int_T \{-\nabla \cdot \underline{a}(x, u_h, \nabla u_h) - b(x, u_h, \nabla u_h)\}\varphi \\ &+ \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E [n \cdot \underline{a}(x, u_h, \nabla u_h)]_E \varphi\end{aligned} \tag{6.4}$$

and

$$\begin{aligned}\langle \widetilde{F}_h(u_h), \varphi\rangle = &\sum_{T \in T_h} \int_T \{-\nabla \cdot \underline{a}_h(x, u_h, \nabla u_h) - b_h(x, u_h, \nabla u_h)\}\varphi \\ &+ \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E [n \cdot \underline{a}_h(x, u_h, \nabla u_h)]_E \varphi.\end{aligned} \tag{6.5}$$

Lemma 5.1, inequalities (5.1), (5.2), the definition of $\widetilde{Y}_h$, and equalities (6.4), (6.5) then imply that

(6.6)

$$\begin{aligned}&\|(\mathrm{Id}_Y - R_h)^*[F(u_h) - \widetilde{F}_h(u_h)]\|_{Y^*} \\ &= \sup_{\substack{\varphi \in Y \\ \|\varphi\|_Y = 1}} \sum_{T \in \mathcal{T}_h} \int_T \{-\nabla \cdot (\underline{a}(x, u_h, \nabla u_h) - \underline{a}_h(x, u_h, \nabla u_h)) \\ &\qquad\qquad - (b(x, u_h, \nabla u_h) - b_h(x, u_h, \nabla u_h))\}\{\varphi - I_h\varphi\} \\ &\qquad + \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E [n \cdot (\underline{a}(x, u_h, \nabla u_h) - \underline{a}_h(x, u_h, \nabla u_h))]_E \{\varphi - I_h \varphi\} \\ &\le c \left\{ \sum_{T \in \mathcal{T}_h} \varepsilon_T^p \right\}^{1/p}\end{aligned}$$

and
(6.7)

$$\begin{aligned}
&\|F(u_h) - \widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*} \\
&\quad = \sup_{\substack{\varphi_h \in \widetilde{Y}_h \\ \|\varphi_h\|_Y = 1}} \sum_{T \in \mathcal{T}_h} \int_T \{-\nabla \cdot (\underline{a}(x, u_h, \nabla u_h) - \underline{a}_h(x, u_h, \nabla u_h)) \\
&\qquad\qquad - (b(x, u_h, \nabla u_h) - b_h(x, u_h, \nabla u_h))\}\varphi_h \\
&\qquad + \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E [n \cdot (\underline{a}(x, u_h, \nabla u_h) - \underline{a}_h(x, u_h, \nabla u_h))]_E \varphi_h \\
&\quad \le c \left\{ \sum_{T \in \mathcal{T}_h} \varepsilon_T^p \right\}^{1/p} .
\end{aligned}$$

Similarly, we obtain
(6.8)

$$\begin{aligned}
&\|(\mathrm{Id}_Y - R_h)^* \widetilde{F}_h(u_h)\|_{Y^*} \\
&\quad = \sup_{\substack{\varphi \in Y \\ \|\varphi\|_Y = 1}} \sum_{T \in \mathcal{T}_h} \int_T \{-\nabla \cdot \underline{a}_h(x, u_h, \nabla u_h) - b_h(x, u_h, \nabla u_h)\}\{\varphi - I_h\varphi\} \\
&\qquad + \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E [n \cdot \underline{a}_h(x, u_h, \nabla u_h)]_E \{\varphi - I_h\varphi\} \\
&\quad \le c \left\{ \sum_{T \in \mathcal{T}_h} \eta_T^p \right\}^{1/p}
\end{aligned}$$

and
(6.9)

$$\begin{aligned}
\|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*} &= \sup_{\substack{\varphi_h \in \widetilde{Y}_h \\ \|\varphi_h\|_Y = 1}} \sum_{T \in \mathcal{T}_h} \int_T \{-\nabla \cdot \underline{a}_h(x, u_h, \nabla u_h) - b_h(x, u_h, \nabla u_h)\}\varphi_h \\
&\quad + \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E [n \cdot \underline{a}_h(x, u_h, \nabla u_h)]_E \varphi_h \\
&\le c \left\{ \sum_{T \in \mathcal{T}_h} \eta_T^p \right\}^{1/p} .
\end{aligned}$$

In order to prove inequality (4.1), consider an arbitrary simplex $T \in \mathcal{T}_h$ and an arbitrary face $E \in \mathcal{E}_{h,\Omega}$ of T and denote by $\widetilde{Y}_h|_\omega$, $\omega \in \{T, \omega_E, \omega_T\}$, the set of all functions $\varphi \in \widetilde{Y}_h$ with $\mathrm{supp}(\varphi) \subset \omega$. Lemma 5.1, equation (6.5), and the definition of $\widetilde{Y}_h$ then yield

$$
\begin{aligned}
c_1 c_4^{-1} h_T \| -\nabla \cdot \underline{a}_h(\cdot, u_h, \nabla u_h) - b_h(\cdot, u_h, \nabla u_h)\|_{0,p;T}
&\le \sup_{v \in \Pi_{m\,|\,T} \setminus \{0\}} \|\nabla(\psi_T v)\|_{0,q;T}^{-1} \\
&\qquad \cdot \int_T \{-\nabla \cdot \underline{a}_h(x, u_h, \nabla u_h) - b_h(x, u_h, \nabla u_h)\} \psi_T v \\
&= \sup_{v \in \Pi_{m\,|\,T} \setminus \{0\}} \|\nabla(\psi_T v)\|_{0,q;T}^{-1} \langle \widetilde{F}_h(u_h), \psi_T v\rangle \\
&\le \sup_{\substack{\varphi \in \widetilde{Y}_{h\,|\,T} \\ \|\varphi\|_Y = 1}} \langle \widetilde{F}_h(u_h), \varphi\rangle
\end{aligned}
\tag{6.10}
$$

and, using inequality (6.10),

$$
\begin{aligned}
& c_2 c_6^{-1} c_7^{-1} h_E^{1/p} \| [n \cdot \underline{a}_h(\cdot, u_h, \nabla u_h)]_E \|_{p;E} \\
&\quad \le \sup_{\sigma \in \Pi_{k\,|\,E} \setminus \{0\}} c_6^{-1} c_7^{-1} h_E^{1/p} \|\sigma\|_{q;E}^{-1} \int_E [n \cdot \underline{a}_h(x, u_h, \nabla u_h)]_E \psi_E P\sigma \\
&\quad = \sup_{\sigma \in \Pi_{k\,|\,E} \setminus \{0\}} c_6^{-1} c_7^{-1} h_E^{1/p} \|\sigma\|_{q;E}^{-1} \\
&\qquad\quad \cdot \left\{ \langle \widetilde{F}_h(u_h), \psi_E P\sigma\rangle \right. \\
&\qquad\qquad \left. - \int_{\omega_E} \{-\nabla \cdot \underline{a}_h(x, u_h, \nabla u_h) - b_h(x, u_h, \nabla u_h)\} \psi_E P\sigma \right\} \\
&\quad \le \sup_{\substack{\varphi \in \widetilde{Y}_{h\,|\,\omega_E} \\ \|\varphi\|_Y = 1}} \langle \widetilde{F}_h(u_h), \varphi\rangle \\
&\qquad\quad + c_6^{-1} h_E \| -\nabla \cdot \underline{a}_h(\cdot, u_h, \nabla u_h) - b_h(\cdot, u_h, \nabla u_h)\|_{0,p;\omega_E} \\
&\quad \le c \sup_{\substack{\varphi \in \widetilde{Y}_{h\,|\,\omega_E} \\ \|\varphi\|_Y = 1}} \langle \widetilde{F}_h(u_h), \varphi\rangle .
\end{aligned}
\tag{6.11}
$$

Inequalities (6.10) and (6.11) imply that

$$
\eta_T \le c \sup_{\substack{\varphi \in \widetilde{Y}_{h\,|\,\omega_T} \\ \|\varphi\|_Y = 1}} \langle \widetilde{F}_h(u_h), \varphi\rangle
\tag{6.12}
$$

and

$$
\left\{ \sum_{T \in \mathcal{T}_h} \eta_T^p \right\}^{1/p} \le c \|\widetilde{F}_h(u_h)\|_{\widetilde{Y}_h^*} .
\tag{6.13}
$$

Inequalities (6.8) and (6.13), in particular, prove inequality (4.1).

Propositions 2.1 and 4.1 and inequalities (6.6), (6.7), (6.8), (6.9), (6.12), and (6.13) yield the following a posteriori error estimates for problem (6.1).

Proposition 6.1. *Let $u \in X$ be a weak solution of problem* (6.1) *which is regular in the sense of Proposition* 2.1, *and let $u_h \in X_h$ be an approximate solution of* *the corresponding discrete problem which is sufficiently close to* u *in the sense of Proposition* 2.1. *Then the following a posteriori error estimates hold*:

$$\|u-u_h\|_{1,r} \le c_1 \left\{ \sum_{T\in\mathcal{T}_h} \eta_T^p \right\}^{1/p} + c_2 \left\{ \sum_{T\in\mathcal{T}_h} \varepsilon_T^p \right\}^{1/p} \\ + c_3\|F(u_h)-F_h(u_h)\|_{Y_h^*} + c_4\|F_h(u_h)\|_{Y_h^*}$$

and

$$\eta_T \le c_5\|u-u_h\|_{1,r;\omega_T} + c_6 \left\{ \sum_{T'\subset\omega_T} \varepsilon_{T'}^p \right\}^{1/p} \quad \forall T\in\mathcal{T}_h.$$

Here, ε_T *and* η_T *are given by equations* (6.2) *and* (6.3), *and* $\|F(u_h)-F_h(u_h)\|_{Y_h^*}$ *and* $\|F_h(u_h)\|_{Y_h^*}$ *are the consistency error of the discretization and the residual of the discrete problem, respectively.*

Remark 6.2. Proposition 6.1 can easily be extended to the case of Neumann boundary conditions. One only has to replace Γ in equations (6.2) and (6.3) by the part of the boundary on which Dirichlet boundary conditions are imposed. The first estimate of Proposition 6.1 also holds if η_T is defined using the original functions $\underline{a}$ and b instead of the projected ones $\underline{a}_h$ and b_h. The ε_T-term then of course disappears. If the functions a and b are sufficiently smooth, one may also use higher-order approximations $\underline{a}_h$ and b_h instead of the present low-order ones. □

As mentioned before, Proposition 6.1 can be applied to example (2) only in the case $\alpha \ge 2$. Observing that for $1<\alpha<2$ the strong monotonicity of F implies the unique solvability of the corresponding weak problem, we obtain from Remark 2.3 and inequalities (6.6), (6.7), (6.8), (6.9), and (6.13) the following result which complements the results of [9].

Proposition 6.3. *Let* $1<\alpha<2$ *and denote by* $u\in W^{1,\alpha}(\Omega)$, $u=0$ *on* Γ, *the unique solution of*

$$\int_\Omega \|\nabla u\|^{\alpha-2}\nabla u\nabla v = \int_\Omega fv \quad \forall v\in W^{1,\alpha}(\Omega),\ v=0 \text{ on } \Gamma.$$

Let $u_h\in X_h$ *be an approximate solution of a discretization of the above problem. Then the following a posteriori error estimates hold*:

$$\|u-u_h\|_{1,\alpha} \le c_1 \left\{ \sum_{T\in\mathcal{T}_h} \eta_T^\alpha \right\}^{1/\alpha} + c_2 \left\{ \sum_{T\in\mathcal{T}_h} \varepsilon_T^\alpha \right\}^{1/\alpha} \\ + c_3\|F(u_h)-F_h(u_h)\|_{Y_h^*} + c_4\|F_h(u_h)\|_{Y_h^*}$$

and

$$\left\{ \sum_{T\in\mathcal{T}_h} \eta_T^\alpha \right\}^{1/\alpha} \le c_5\|u-u_h\|_{1,\alpha}^{1/(\alpha-1)} + c_6 \left\{ \sum_{T\in\mathcal{T}_h} \varepsilon_T^\alpha \right\}^{1/\alpha(\alpha-1)} .$$

Here, ε_T, η_T, $\|F(u_h)-F_h(u_h)\|_{Y_h^*}$, *and* $\|F_h(u_h)\|_{Y_h^*}$ *are as in Proposition* 6.1. *Moreover,*

$$\varepsilon_T = h_T\|f-\Pi_{0,T}f\|_{0,\alpha;T} \quad \forall T\in\mathcal{T}_h,$$

if u_h *is piecewise linear.*

As mentioned before, example (6) exhibits a simple bifurcation from the trivial branch at the simple eigenvalues of the Laplacian. Combining the results of §3 with those of this section, we obtain the following a posteriori error estimate.

Proposition 6.4. *Denote by $\lambda^* \in \mathbb{R}$ and $u^* \in W^{1,p}(\Omega)$, $u^* = 0$ on Γ, $p > n$, a simple eigenvalue of the Laplace equation with homogeneous Dirichlet boundary conditions and a corresponding eigenfunction with $\int_\Omega u^* = 1$. Let $\lambda_h \in \mathbb{R}$ and $u_h \in X_h$ be a solution of*

$$\int_\Omega \nabla u_h \nabla v_h - \lambda_h \int_\Omega u_h v_h + \int_\Omega u_h^\beta v_h = 0 \quad \forall v_h \in X_h,$$

where $X_h \subset \{v \in W^{1,p}(\Omega) \cap W^{1,\infty}(\Omega) : v = 0 \text{ on } \Gamma\}$ is a finite element space corresponding to $\mathcal{T}_h$ consisting of piecewise polynomials, and where $\beta \in \mathbb{N}$, $\beta \geq n$. If λ_h and u_h are sufficiently close to λ^* and u^*, the following a posteriori error estimates hold:

$$|\lambda_h - \lambda^*| + \|u_h - u^*\|_{1,p} \leq c_1 \left\{ \left| \int_\Omega u_h - 1 \right| + \left\{ \sum_{T \in \mathcal{T}_h} \eta_T^p \right\}^{1/p} \right\}$$

and

$$\left| \int_\Omega u_h - 1 \right| + \left\{ \sum_{T \in \mathcal{T}_h} \eta_T^p \right\}^{1/p} \leq c_2 \{ |\lambda_h - \lambda^*| + \|u_h - u^*\|_{1,p} \},$$

where

$$\eta_T := \left\{ h_T^p \| -\Delta u_h - \lambda_h u_h \|_{0,p;T}^p + \sum_{E \subset \partial T \setminus \Gamma} h_E \| [\partial_n u_h]_E \|_{p;E}^p \right\}^{1/p}.$$

Proof. Observe that the consistency error of the above discretization vanishes; Proposition 6.4 then follows from inequalities (6.6), (6.7), (6.8), and (6.9) and the results of the third part of §3 with $l \in \mathcal{L}(V, \mathbb{R})$ given by

$$l(v) := \int_\Omega v \quad \forall v \in W^{1,p}(\Omega), \; v = 0 \text{ on } \Gamma. \quad \square$$

When comparing Propositions 6.1 and 6.4, we remark that the latter only yields global lower bounds on the error. This is due to the global nature of the functional l defined above.

We conclude this section with a simple example of an a posteriori error estimator which is based on the solution of auxiliary local problems and which generalizes the estimator introduced in [4, 5]. For simplicity we assume that $p = q = r = 2$. We choose an arbitrary vertex x_0 in the partition $\mathcal{T}_h$ and keep it fixed in what follows. Denote by $\mathcal{T}_0$ and $\mathcal{E}_0$ the set of all $T \in \mathcal{T}_h$ and of all $E \in \mathcal{E}_h$, respectively, which have x_0 as a vertex. Put $\omega_0 := \bigcup_{T \in \mathcal{T}_0} T$. Let

$$\widehat{X}_h := \widehat{Y}_h := \widetilde{Y}_h|_{\omega_0},$$

and define the operator $B \in \mathcal{L}(\widehat{X}_h, \widehat{Y}_h^*)$ by

$$\langle Bu, \varphi \rangle := \int_{\omega_0} \nabla \varphi^t A_0 \nabla u \quad \forall u \in \widehat{X}_h, \; \varphi \in \widehat{Y}_h,$$

where

$$A_0 := A(x_0, u_h(x_0), \pi_{0,\omega_0}(\nabla u_h)).$$

Note that the operator B is obtained by first linearizing around u_h the differential operator associated with problem (6.1), then freezing the coefficients of the resulting linear operator at x_0, and finally retaining only the principal part of the linear constant-coefficient operator. Since ∇u_h may be discontinuous, its value at x_0 is approximated by the L^2-projection $\pi_{0,\omega_0}(\nabla u_h)$. Other constructions are of course also possible.

Since the matrix $A(x, y, z)$ is symmetric and positive definite for all $x \in \Omega$, $y \in \mathbb{R}$, $z \in \mathbb{R}^n$, and since the functions in $\widehat{X}_h = \widehat{Y}_h$ vanish on $\partial\omega_0$, we immediately obtain from Korn's inequality that $B \in \operatorname{Isom}(\widehat{X}_h, \widehat{Y}_h^*)$. Let $u_0 \in \widehat{X}_h$ be the unique solution of

$$\langle Bu_0, \psi\rangle = \langle \widetilde{F}_h(u_h), \varphi\rangle \quad \forall\varphi \in \widehat{Y}_h, \tag{6.14}$$

and set

$$\eta_{x_0} := \|u_0\|_{1,2;\omega_0}. \tag{6.15}$$

Note that problem (6.14) is equivalent to

$$\int_{\omega_0} \nabla\varphi^t A_0 \nabla u_0 = \int_{\omega_0} \underline{a}_h(x, u_h, \nabla u_h)\nabla\varphi - \int_{\omega_0} b_h(x, u_h, \nabla u_h)\varphi \quad \forall\varphi \in \widehat{Y}_h.$$

This shows that η_{x_0} falls into the class of a posteriori error estimators originally introduced in [4, 5] for the Poisson equation.

Lemma 5.1 and equations (6.5) and (6.12) immediately imply that

$$\underline{c}\|\widetilde{F}_h(u_h)\|_{\widehat{Y}_h} \le \left\{\sum_{T\in\mathcal{T}_0} \eta_T^2\right\}^{1/2} \le \bar{c}\|\widetilde{F}_h(u_h)\|_{\widehat{Y}_h}.$$

Together with Proposition 4.3, this yields the following result.

Proposition 6.5. *Let x_0 be an arbitrary vertex in the triangulation $\mathcal{T}_h$. Then there are two constants c_1, c_2, which only depend on the polynomial degree of the space X_h and on the ratio h_T/ϱ_T, such that the following inequalities hold:*

$$c_1\left\{\sum_{T\in\mathcal{T}_0} \eta_T^2\right\}^{1/2} \le \eta_{x_0} \le c_2\left\{\sum_{T\in\mathcal{T}_0} \eta_T^2\right\}^{1/2}.$$

Here, η_T and η_{x_0} are given by equations (6.3) *and* (6.15), *respectively.*

7. Eigenvalue problems for scalar linear elliptic operators of 2nd order

As an example for the treatment of eigenvalue problems, we consider in this section the problem

$$\begin{aligned} -\nabla\cdot(A(x)\nabla u) + d(x)u &= \lambda u \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{7.1}$$

Here, $d \in C(\Omega, \mathbb{R}_+)$ and $A \in C^1(\Omega, \mathbb{R}^{n\times n})$ are such that A is symmetric and uniformly positive definite on Ω. Of course, we are only interested in solutions u which do not identically vanish on Ω.

When considering λ as a parameter, problem (7.1) can be treated as a bifurcation problem similar to example (6) of the previous section. Here, we adopt a different strategy and define

$$X := Y := \mathbb{R} \times \{u \in W^{1,2}(\Omega) : u = 0 \text{ on } \Gamma\},$$
$$\|\cdot\|_X := \|\cdot\|_Y := \{|\cdot|^2 + \|\cdot\|_{1,2}^2\}^{1/2},$$
$$\langle F([\lambda, u]), [\mu, v]\rangle := \int_\Omega \{\nabla v^t A \nabla u + duv - \lambda uv\} + \mu\left\{\int_\Omega u^2 - 1\right\}.$$

Then, $[\lambda, u] \in X$, $\|u\|_{0,2} = 1$, is a weak solution of problem (7.1) if and only if it is a solution of equation (1.1). Moreover, one easily checks that $[\lambda, u]$ is a regular solution in the sense of Proposition 2.1 if and only if λ is a simple eigenvalue of the differential operator associated with problem (7.1).

As in the previous section, we do not specify the discretization of problem (7.1) in detail. We only assume that

$$X_h = \mathbb{R} \times V_h \subset X, \qquad Y_h = \mathbb{R} \times W_h \subset Y,$$
$$\langle F_h([\lambda_h, u_h]), [\mu_h, v_h]\rangle = \langle F([\lambda_h, u_h]), [\mu_h, v_h]\rangle \quad \forall [\lambda_h, u_h] \in X_h, \ [\mu_h, v_h] \in Y_h,$$

where V_h, W_h are finite element spaces corresponding to $\mathcal{T}_h$ which consist of affinely equivalent elements in the sense of [10] and which satisfy $\{v_h \in S_h^{1,0} : v_h = 0 \text{ on } \Gamma\} \subset W_h$. Obviously, the consistency error of the above discretization vanishes. Moreover, $[\lambda_h, u_h] \in X_h$ is a solution of equation (1.2) if and only if

$$\int_\Omega \{\nabla v_h^t A \nabla u_h + d u_h v_h\} = \lambda_h \int_\Omega u_h v_h \quad \forall v_h \in W_h, \qquad \int_\Omega u_h^2 = 1. \tag{7.2}$$

Hence, problem (1.2) is equivalent to a standard finite-dimensional eigenvalue problem. In what follows, we will always assume that $[\lambda_h, u_h] \in X_h$ is a solution of problem (7.2).

Let m be the maximal polynomial degree of the functions in W_h. Proceeding as in the previous section, we set

(7.3)
$$A_h := \sum_{T \in \mathcal{T}_h} \pi_{1,T} A,$$
$$d_h := \sum_{T \in \mathcal{T}_h} \pi_{0,T} d,$$
$$R_h := [0, I_h],$$
$$\widetilde{Y}_h := \mathbb{R} \times \operatorname{span}\{\psi_T v, \psi_E P\sigma : v \in \Pi_m|_T, \ \sigma \in \Pi_m|_E, \ T \in \mathcal{T}_h, \ E \in \mathcal{E}_{h,\Omega}\},$$
$$\varepsilon_T := \left\{ h_T^2 \| - \nabla \cdot ((A - A_h)\nabla u_h) + (d - d_h) u_h\|_{0,2;T}^2 + \sum_{E \subset \partial T \setminus \Gamma} h_E \|[n \cdot ((A - A_h)\nabla u_h)]_E\|_{2;E}^2 \right\}^{1/2},$$

$$(7.4)\qquad \eta_T := \left\{ h_T^2 \| -\nabla\cdot(A_h\nabla u_h) + d_h u_h - \lambda_h u_h\|_{0,2;T}^2 + \sum_{E\subset\partial T\backslash\Gamma} h_E \|[n\cdot(A_h\nabla u_h)]_E\|_{2;E}^2 \right\}^{1/2},$$

and define $\widetilde{F}_h$ in the same way as F with A and d replaced by A_h and d_h, respectively. Note that

$$(7.5)\qquad \varepsilon_T \le c h_T^2\{\|A\|_{2,\infty;T}\|u_h\|_{1,2;T} + \|d\|_{1,\infty;T}\|u_h\|_{0,2;T}\}.$$

With the same arguments as in the previous section we conclude that

$$\begin{aligned}\langle F([\lambda_h, u_h]), [\mu, v]\rangle &= \sum_{T\in\mathcal{T}_h}\int_T \{-\nabla\cdot(A\nabla u_h) + du_h - \lambda_h u_h\}v \\ &\quad + \sum_{E\in\mathcal{E}_{h,\Omega}}\int_E [n\cdot(A\nabla u_h)]_E v \quad \forall[\mu, v]\in Y, \\ \langle \widetilde{F}_h([\lambda_h, u_h]), [\mu, v]\rangle &= \sum_{T\in\mathcal{T}_h}\int_T \{-\nabla\cdot(A_h\nabla u_h) + d_h u_h - \lambda_h u_h\}v \\ &\quad + \sum_{E\in\mathcal{E}_{h,\Omega}}\int_E [n\cdot(A_h\nabla u_h)]_E v \quad \forall[\mu, v]\in Y,\end{aligned}$$

$$(7.6)\qquad \|(\mathrm{Id}_Y - R_h)^*[F([\lambda_h, u_h]) - \widetilde{F}_h([\lambda_h, u_h])]\|_{Y^*} \le c\left\{\sum_{T\in\mathcal{T}_h}\varepsilon_T^2\right\}^{1/2},$$

$$(7.7)\qquad \|F([\lambda_h, u_h]) - \widetilde{F}_h([\lambda_h, u_h])\|_{\widetilde{Y}_h^*} \le c\left\{\sum_{T\in\mathcal{T}_h}\varepsilon_T^2\right\}^{1/2},$$

$$(7.8)\qquad \|(\mathrm{Id}_Y - R_h)^*\widetilde{F}_h([\lambda_h, u_h])\|_{Y^*} \le c\left\{\sum_{T\in\mathcal{T}_h}\eta_T^2\right\}^{1/2},$$

$$(7.9)\qquad \|\widetilde{F}_h([\lambda_h, u_h])\|_{\widetilde{Y}_h^*} \le c\left\{\sum_{T\in\mathcal{T}_h}\eta_T^2\right\}^{1/2},$$

$$(7.10)\qquad \eta_T \le \sup_{\substack{[0,v]\in\widetilde{Y}_h \\ \operatorname{supp} v\subset\omega_T}} \|[0, v]\|_Y^{-1}\langle\widetilde{F}_h([\lambda_h, u_h]), [0, v]\rangle.$$

Inequalities (7.8) and (7.10), in particular, prove inequality (4.1).

Propositions 2.1 and 4.1 and inequalities (7.6)–(7.10) yield the following a posteriori error estimate for problem (7.1).

Proposition 7.1. *Let λ be a simple eigenvalue of the differential operator associated with problem* (7.1), *and let u be a corresponding eigenfunction with*

$\|u\|_{0,2} = 1$. *Let* $[\lambda_h, u_h] \in X_h$ *be a solution of problem* (7.2) *which is sufficiently close to* $[\lambda, u]$ *in the sense of Proposition* 2.1. *Then the following a posteriori error estimates hold*:

$$|\lambda - \lambda_h| + \|u - u_h\|_{1,2} \leq c_1 \left\{ \sum_{T \in \mathcal{T}_h} \eta_T^2 \right\}^{1/2} + c_2 \left\{ \sum_{T \in \mathcal{T}_h} \varepsilon_T^2 \right\}^{1/2}$$

and

$$\left\{ \sum_{T \in \mathcal{T}_h} \eta_T^2 \right\}^{1/2} \leq c_3 \{ |\lambda - \lambda_h| + \|u - u_h\|_{1,2} \} + c_4 \left\{ \sum_{T \in \mathcal{T}_h} \varepsilon_T^2 \right\}^{1/2},$$

where the constants $c_1, \dots, c_4$ *only depend on the polynomial degree of the spaces* V_h *and* W_h *and on the ratio* h_T/ϱ_T, *and where* ε_T *and* η_T *are given by equations* (7.3) *and* (7.4), *respectively.*

Remark 7.2. The condition that $[\lambda_h, u_h]$ has to be sufficiently close to $[\lambda, u]$ essentially means that $|\lambda - \lambda_h|$ has to be smaller than the distance of λ to its neighboring eigenvalues. In contrast to Proposition 6.1, we obtain in Proposition 7.1 only a global lower bound on the error. This is due to the global nature of the constraint $\int_\Omega u^2 = 1$ inherent in the definition of F. Proposition 7.1 can easily be extended to the case of Neumann boundary conditions. One only has to replace Γ in equations (7.3) and (7.4) by the part of the boundary on which Dirichlet boundary conditions are imposed. □

8. Stationary, incompressible Navier-Stokes equations

As an example for the treatment of elliptic systems we consider the stationary, incompressible Navier-Stokes equations

$$(8.1) \qquad \begin{aligned} -\nu \Delta \mathbf{u} + (\mathbf{u} \cdot \nabla)\mathbf{u} + \nabla p &= \mathbf{f} \quad \text{in } \Omega, \\ \nabla \cdot \mathbf{u} &= 0 \quad \text{in } \Omega, \\ \mathbf{u} &= 0 \quad \text{on } \Gamma, \end{aligned}$$

where $\nu > 0$ is the constant viscosity of the fluid.

In order to cast problem (8.1) into the framework of §2, set

$$M := \{\mathbf{u} \in W^{1,2}(\Omega)^n : \mathbf{u} = 0 \text{ on } \Gamma\}, \qquad Q := \left\{ p \in L^2(\Omega) : \int_\Omega p = 0 \right\},$$

and define

$$X := Y := M \times Q, \qquad \|\cdot\|_X := \|\cdot\|_Y := \{\|\cdot\|_{1,2}^2 + \|\cdot\|_{0,2}^2\}^{1/2},$$

$$\langle F([\mathbf{u}, p]), [\mathbf{v}, q] \rangle := \nu \int_\Omega \nabla \mathbf{u} \nabla \mathbf{v} + \int_\Omega (\mathbf{u} \cdot \nabla)\mathbf{u}\mathbf{v} - \int_\Omega p \nabla \cdot \mathbf{v} + \int_\Omega q \nabla \cdot \mathbf{u} - \int_\Omega \mathbf{f}\mathbf{v}.$$

Let $M_h \subset M$ and $Q_h \subset Q$ be two finite element spaces corresponding to $\mathcal{T}_h$ consisting of affinely equivalent elements in the sense of [10]. We assume that there are two integers $k, l \geq 1$ such that

$$[S_h^{1,0}]^n \cap M \subset M_h \subset [S_h^{k,0}]^n$$

and

$$S_h^{1,0} \cap Q \subset Q_h \subset S_h^{l,0} \quad \text{or} \quad S_h^{0,-1} \cap Q \subset Q_h \subset S_h^{l,-1}.$$

Define

$$X_h := Y_h := M_h \times Q_h ,$$

$$\begin{aligned}
&\langle F_h([\mathbf{u}_h , p_h]) , [\mathbf{v}_h , q_h]\rangle \\
&\quad := \langle F([\mathbf{u}_h , p_h]) , [\mathbf{v}_h , q_h]\rangle \\
&\qquad + \delta \sum_{T\in\mathcal{T}_h} h_T^2 \int_T \{-\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \mathbf{f}\}\{(\mathbf{u}_h\cdot\nabla)\mathbf{v}_h + \nabla q_h\} \\
&\qquad + \delta \sum_{E\in\mathcal{E}_{h,\Omega}} h_E \int_E [p_h]_E[q_h]_E + \alpha\delta \int_\Omega \nabla\cdot\mathbf{u}_h\nabla\cdot\mathbf{v}_h .
\end{aligned} \tag{8.2}$$

Here, $\alpha \geq 0$, $\delta \geq 0$ are stability parameters. If $\alpha > 0$, $\delta > 0$, the above discretization is capable of stabilizing both the influence of the convection term and of the divergence constraint without any conditions about the spaces M_h, Q_h or the Peclet number $h_T\nu^{-1}$ (cf. [23], where also optimal a priori error estimates are established). The case $\alpha = \delta = 0$ corresponds to the standard mixed finite element discretization of problem (8.1). The spaces M_h, Q_h then have to satisfy the Babuška-Brezzi condition

$$\inf_{p_h\in Q_h\setminus\{0\}} \sup_{\mathbf{u}_h\in M_h\setminus\{0\}} \frac{\int_T p_h\nabla\cdot\mathbf{u}_h}{\|p_h\|_{0,2}\|\mathbf{u}_h\|_{1,2}} \geq \beta > 0 \tag{8.3}$$

with a constant β independent of h. Moreover, the Peclet number $h_T\nu^{-1}$ must be sufficiently small in order to balance the influence of the convection term (cf. [17], where examples of spaces M_h, Q_h satisfying inequality (8.3) are also given).

If $\alpha = \delta = 0$, the consistency error obviously vanishes. If $\delta > 0$, we conclude from standard inverse estimates that it is bounded by

(8.4)

$$\begin{aligned}
&\|F([\mathbf{u}_h , p_h]) - F_h([\mathbf{u}_h , p_h])\|_{Y_h^*} \\
&\quad = \sup_{\substack{[\mathbf{v}_h , q_h]\in Y_h \\ \|[\mathbf{v}_h , q_h]\|_Y=1}} \delta \sum_{T\in\mathcal{T}_h} h_T^2 \int_T \{-\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \mathbf{f}\} \\
&\qquad\qquad \cdot\{(\mathbf{u}_h\cdot\nabla)\mathbf{v}_h + \nabla q_h\} \\
&\qquad + \delta \sum_{E\in\mathcal{E}_h , \Omega} h_E \int_E [p_h]_E[q_h]_E + \alpha\delta \int_\Omega \nabla\cdot\mathbf{u}_h\nabla\cdot\mathbf{v}_h \\
&\quad \leq c(1+\alpha)\delta(1+\|\mathbf{u}_h\|_{1,2}) \\
&\qquad \cdot\left\{ \sum_{T\in\mathcal{T}_h} [h_T^2\| - \nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \mathbf{f}\|_{0,2;T}^2 + \|\nabla\cdot\mathbf{u}_h\|_{0,2;T}^2] \right. \\
&\qquad\qquad \left. + \sum_{E\in\mathcal{E}_{h,\Omega}} h_E\|[p_h]_E\|_{2;E}^2 \right\}^{1/2} .
\end{aligned}$$

In order to cast this discretization into the framework of §4, we define $\widetilde{F}_h$ in the same way as F with $\mathbf{f}$ replaced by $\pi_{0,T}\mathbf{f}$ and set

$$\begin{gathered} R_h[\mathbf{u}, p] := [I_h u_1, \dots, I_h u_n, 0], \\ \widetilde{Y}_h := \operatorname{span}\{[\psi_T \mathbf{v}, 0], [\psi_E P\sigma, 0], [0, \psi_T p] : \mathbf{v} \in [\Pi_m|_T]^n, \ \sigma \in [\Pi_{m'}|_E]^n, \\ p \in \Pi_{k-1}|_T, \ T \in \mathcal{T}_h, \ E \in \mathcal{E}_{h,\Omega}\}, \end{gathered}$$

where $m := \max\{2k-1, l-1\}$ and $m' := \max\{k-1, l\}$.

Lemma 5.1 and inequality (5.1) immediately imply

$$\tag{8.5} \begin{aligned} \|\widetilde{F}_h([\mathbf{u}_h, p_h]) - F([\mathbf{u}_h, p_h])\|_{\widetilde{Y}_h^*} &= \sup_{\substack{[\mathbf{v}_h, q_h] \in \widetilde{Y}_h \\ \|[\mathbf{v}_h, q_h]\|_Y = 1}} \sum_{T \in \mathcal{T}_h} \int_T (\mathbf{f} - \pi_{0,T}\mathbf{f})\mathbf{v}_h \\ &\le c \left\{ \sum_{T \in \mathcal{T}_h} h_T^2 \|\mathbf{f} - \pi_{0,T}\mathbf{f}\|_{0,2;T}^2 \right\}^{1/2} \end{aligned}$$

and

$$\tag{8.6} \begin{aligned} &\|(\mathrm{Id}_Y - R_h)^*[\widetilde{F}_h([\mathbf{u}_h, p_h]) - F([\mathbf{u}_h, p_h])\|_{Y^*} \\ &\quad = \sup_{\substack{[\mathbf{v}, q] \in Y \\ \|[\mathbf{v}, q]\|_Y = 1}} \sum_{T \in \mathcal{T}_h} \sum_{i=1}^n \int_T (f_i - \pi_{0,T} f_i)(v_i - I_h v_i) \\ &\quad \le c \left\{ \sum_{T \in \mathcal{T}_h} h_T^2 \|\mathbf{f} - \pi_{0,T}\mathbf{f}\|_{0,2;T}^2 \right\}^{1/2}. \end{aligned}$$

For abbreviation, we define for all $T \in \mathcal{T}_h$

$$\tag{8.7} \begin{aligned} \eta_T := \Bigg\{ & h_T^2 \| - \nu \Delta \mathbf{u}_h + (\mathbf{u}_h \cdot \nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\|_{0,2;T}^2 \\ & + \sum_{E \subset \partial T \setminus \Gamma} h_E \|[\nu \partial_n \mathbf{u}_h - p_h \mathbf{n}]_E\|_{2;E}^2 + \|\nabla \cdot \mathbf{u}_h\|_{0,2;T}^2 \Bigg\}^{1/2}. \end{aligned}$$

Observing that the identity

$$\tag{8.8} \begin{aligned} &\langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\mathbf{v}, q] \rangle \\ &\quad = \sum_{T \in \mathcal{T}_h} \left\{ \int_T \{-\nu \Delta \mathbf{u}_h + (\mathbf{u}_h \cdot \nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\}\mathbf{v} + \int_T q \nabla \cdot \mathbf{u}_h \right\} \\ &\quad\quad + \sum_{E \in \mathcal{E}_{h,\Omega}} \int_E [\nu \partial_n \mathbf{u}_h - p_h \mathbf{n}]_E \mathbf{v} \end{aligned}$$

holds for all $[\mathbf{v}, q] \in Y$, we conclude from Lemma 5.1 and inequalities (5.1), (5.2) that

(8.9)
$$\begin{aligned}
&\|\widetilde{F}_h([\mathbf{u}_h, p_h])\|_{\widetilde{Y}_h^*} \\
&\quad= \sup_{\substack{[\mathbf{v}_h, q_h]\in\widetilde{Y}_h \\ \|[\mathbf{v}_h, q_h]\|_Y=1}} \sum_{T\in\mathcal{T}_h}\left\{\int_T\{-\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\}\mathbf{v}_h + \int_T q_h\nabla\cdot\mathbf{u}_h\right\} \\
&\qquad\qquad + \sum_{E\in\mathcal{E}_{h,\Omega}}\int_E[\nu\partial_n\mathbf{u}_h - p_h\mathbf{n}]_E\mathbf{v}_h \\
&\quad\le c\left\{\sum_{T\in\mathcal{T}_h}\eta_T^2\right\}^{1/2}
\end{aligned}$$

and

(8.10)
$$\begin{aligned}
&\|(\mathrm{Id}_Y - R_h)^*\widetilde{F}_h([\mathbf{u}_h, p_h])\|_{Y^*} \\
&\quad= \sup_{\substack{[\mathbf{v}, q]\in Y \\ \|[\mathbf{v}, q]\|_Y=1}} \sum_{T\in\mathcal{T}_h}\left\{\sum_{i=1}^n\int_T\{-\nu\Delta u_{h,i} + (\mathbf{u}_h\cdot\nabla)u_{h,i} + \partial_i p_h - \pi_{0,T}f_i\}\right. \\
&\qquad\qquad\left.\cdot\{v_i - I_h v_i\} + \int_T q\nabla\cdot u_h\right\} \\
&\qquad + \sum_{E\in\mathcal{E}_{h,\Omega}}\sum_{i=1}^n\int_E[\nu\partial_n u_{h,i} - p_h n_i]_E(v_i - I_h v_i) \\
&\quad\le c\left\{\sum_{T\in\mathcal{T}_h}\eta_T^2\right\}^{1/2}.
\end{aligned}$$

In order to prove inequality (4.1), consider an arbitrary simplex $T \in \mathcal{T}_h$ and an arbitrary face $E \in \mathcal{E}_{h,\Omega}$ of T and define $\widetilde{Y}_h|_\omega$, $\omega \in \{T, \omega_E, \omega_T\}$, as in §6. The definition of $\widetilde{Y}_h$, equation (8.8), and Lemma 5.1 then yield the estimates

$$\begin{aligned}
&c_1\|\nabla\cdot\mathbf{u}_h\|_{0,2;T} \\
&\quad\le \sup_{r\in\Pi_{k-1}|_T\setminus\{0\}} \|r\|_{0,2;T}^{-1}\int_T\nabla\cdot u_h\psi_T r \\
&\quad= \sup_{r\in\Pi_{r-1}|_T\setminus\{0\}} \langle\widetilde{F}_h([\mathbf{u}_h, p_h]), [0, \psi_T r]\rangle\|r\|_{0,2;T}^{-1} \qquad (8.11) \\
&\quad\le \sup_{\substack{[\mathbf{v}, q]\in\widetilde{Y}_{h\,|\,T} \\ \|[\mathbf{v}, q]\|_Y=1}} \langle\widetilde{F}_h([\mathbf{u}_h, p_h]), [\mathbf{v}, q]\rangle,
\end{aligned}$$

(8.12)
$$\begin{aligned}
&c_1 c_4^{-1} h_T \| -\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\|_{0,2;T} \\
&\quad\le \sup_{\mathbf{w}\in[\Pi_m|_T]^n\setminus\{0\}} \|\nabla(\psi_T\mathbf{w})\|_{0,2;T}^{-1} \int_T \{-\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\}\psi_T\mathbf{w} \\
&\quad= \sup_{\mathbf{w}\in[\Pi_m|_T]^n\setminus\{0\}} \|\nabla(\psi_T\mathbf{w})\|_{0,2;T}^{-1} \langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\psi_T\mathbf{w}, 0]\rangle \\
&\quad\le \sup_{\substack{[\mathbf{v},q]\in\widetilde{Y}_h|_T \\ \|[\mathbf{v},q]\|_Y=1}} \langle \widetilde{F}_h([\mathbf{u}_h, p_h], [\mathbf{v}, q]\rangle
\end{aligned}$$

and, using inequality (8.12),
(8.13)
$$\begin{aligned}
&c_2 c_6^{-1} c_7^{-1} h_E^{1/2} \|[\nu\partial_n\mathbf{u}_h - p_h\mathbf{n}]_E\|_{2;E} \\
&\quad\le \sup_{\sigma\in[\Pi_{m'}|_E]^n\setminus\{0\}} c_2 c_6^{-1} c_7^{-1} h_E^{1/2} \|\sigma\|_{2;E}^{-1} \int_E [\nu\partial_n\mathbf{u}_h - p_h\mathbf{n}]\psi_E P\sigma \\
&\quad= \sup_{\sigma\in[\Pi_{m'}|_E]^n\setminus\{0\}} c_2 c_6^{-1} c_7^{-1} h_E^{1/2} \|\sigma\|_{2;E}^{-1} \\
&\qquad\qquad \cdot \Bigg\{ \langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\psi_E P\sigma, 0]\rangle \\
&\qquad\qquad\qquad - \int_{\omega_E} \{-\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\}\psi_E P\sigma \Bigg\} \\
&\quad\le \sup_{\substack{[\mathbf{v},q]\in\widetilde{Y}_h|_{\omega_E} \\ \|[\mathbf{v},q]\|_Y=1}} \langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\mathbf{v}, q]\rangle \\
&\qquad\qquad + c_6^{-1} h_E \| -\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\|_{0,2;\omega_E} \\
&\quad\le c \sup_{\substack{[\mathbf{v},q]\in\widetilde{Y}_h|_{\omega_E} \\ \|[\mathbf{v},q]\|_Y=1}} \langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\mathbf{v}, q]\rangle.
\end{aligned}$$

Inequalities (8.11)-(8.13) imply

$$\eta_T \le c \sup_{\substack{[\mathbf{v},q]\in\widetilde{Y}_h|_{\omega_T} \\ \|[\mathbf{v},q]\|_Y=1}} \langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\mathbf{v}, q]\rangle \tag{8.14}$$

and

$$\left\{\sum_{T\in\mathscr{T}_h} \eta_T^2\right\}^{1/2} \le c\|\widetilde{F}_h([\mathbf{u}_h, p_h])\|_{\widetilde{Y}_h^*}. \tag{8.15}$$

Inequalities (8.9), (8.10), and (8.15) prove inequality (4.1) and show that, up to multiplicative constants, $\|\widetilde{F}_h([\mathbf{u}_h, p_h])\|_{\widetilde{Y}_h^*}$ is bounded from above and from below by $\{\sum_T \eta_T^2\}^{1/2}$. Propositions 2.1 and 4.1 and inequalities (8.4), (8.5), (8.6), (8.9), and (8.14) now yield the following a posteriori error estimate, which is a generalization of the results in [25, 27].

Proposition 8.1. *Let* $[\mathbf{u}, p]$ *be a weak solution of problem* (8.1) *which is regular in the sense of Proposition* 2.1, *and let* $[\mathbf{u}_h, p_h] \in X_h$ *be a solution of*

$$\langle F_h([\mathbf{u}_h, p_h]), [\mathbf{v}_h, q_h]\rangle = 0, \quad \forall[\mathbf{v}_h, q_h] \in Y_h,$$

where F_h *is given in equation* (8.2), *which is sufficiently close to* $[\mathbf{u}, p]$ *in the sense of Proposition* 2.1. *Then the following a posteriori error estimates hold*:

$$\begin{aligned}\{\|\mathbf{u}-\mathbf{u}_h\|_{1,2}^2 + \|p-p_h\|_{0,2}^2\}^{1/2} \le c_1[1+(1+\alpha)\delta(1+\|\mathbf{u}_h\|_{1,2})]\left\{\sum_{T\in\mathcal{T}_h}\eta_T^2\right\}^{1/2}&\\ + c_2\left\{\sum_{T\in\mathcal{T}_h} h_T^2\|\mathbf{f}-\pi_{0,T}\mathbf{f}\|_{0,2;T}^2\right\}^{1/2}&,\end{aligned}$$

$$\begin{aligned}\eta_T \le c_3\{\|\mathbf{u}-\mathbf{u}_h\|_{1,2;\omega_T}^2 + \|p-p_h\|_{0,2;\omega_T}^2\}^{1/2}&\\ + c_4\left\{\sum_{T'\subset\omega_T} h_{T'}^2\|\mathbf{f}-\pi_{0,T'}\mathbf{f}\|_{0,2;T'}^2\right\}^{1/2}&,\end{aligned}$$

where η_T *is given by equation* (8.7) *and the constants* $c_1, \dots, c_4$ *only depend on the polynomial degrees of the spaces* M_h, Q_h *and on the ratio* h_T/ϱ_T.

Remark 8.2. Proposition 8.1 can be extended to the case of the slip boundary condition

$$\mathbf{u}\cdot\mathbf{n} = \mathbf{T}(\nu\mathbf{u}, p) - [\mathbf{n}\cdot\mathbf{T}(\nu\mathbf{u}, p)\cdot\mathbf{n}]\mathbf{n} = 0,$$

where

$$\mathbf{T}(\mathbf{u}, p) := \left(\frac{1}{2}(\partial_i u_j + \partial_j u_i) - p\delta_{ij}\right)_{1\le i,j\le n}$$

denotes the stress tensor. One then has to replace $\nu\nabla\mathbf{u} - p\mathbf{I}$ in equation (8.7) by $\mathbf{T}(\nu\mathbf{u}, p)$, and Γ by the part of the boundary on which the no-slip condition $\mathbf{u} = 0$ is imposed. Here, $\mathbf{I} := (\delta_{ij})_{1\le i,j\le n}$ denotes the unit tensor. Of course, the discretization then also has to take account of the different boundary condition (cf., e.g., [24, 26]). □

Remark 8.3. The previous results can also be extended to non-Newtonian fluids. Combining the arguments used to establish Propositions 6.1, 6.3, and 8.1, one can prove that the error estimator of [9] also yields local lower bounds similar to the second estimate of Proposition 8.1. □

Next, we introduce an a posteriori error estimator for problem (8.1), which is based on the solution of discrete local Stokes problems and which fits into the framework of Proposition 4.3. This estimator is an extension to the Navier-Stokes equations of the one introduced in [4, 5] for the Poisson equation.

We choose an arbitrary vertex x_0 in the partition $\mathcal{T}_h$ and keep it fixed in what follows. Let ω_0, $\mathcal{T}_0$ and $\mathcal{E}_0$ be as in §6. Put

$$\begin{aligned}&M_0 := \operatorname{span}\{\psi_T\mathbf{v}, \psi_E P\sigma : \mathbf{v}\in[\Pi_{m''}|_T]^n, \ \sigma\in[\Pi_{m'}|_E]^n, \ T\in\mathcal{T}_0, \ E\in\mathcal{E}_0\},\\ &Q_0 := \operatorname{span}\{\psi_T p : p\in\Pi_{k-1}|_T, \ T\in\mathcal{T}_0\},\end{aligned}$$

where $m := \max\{2k-1, l-1\}$, $m' := \max\{k-1, l\}$, and $m'' := \max\{m, k+n-1\}$, and define

$$\widehat{X}_h := \widehat{Y}_h := M_0 \times Q_0,$$
$$\langle B([\mathbf{v}, q]), [\mathbf{w}, r]\rangle := \nu \int_{\omega_0} \nabla\mathbf{v}\nabla\mathbf{w} - \int_{\omega_0} q\nabla\cdot\mathbf{w} + \int_{\omega_0} r\nabla\cdot\mathbf{v} \quad \forall[\mathbf{v}, q],\ [\mathbf{w}, r] \in \widehat{X}_h.$$

The definition of m'' implies that $\psi_T\nabla q \in M_0$ for all $q \in Q_0$. Together with Lemma 5.1, this shows that the spaces M_0, Q_0 satisfy an analogon of equation (8.3). Hence, we have $B \in \mathrm{Isom}(\widehat{X}_h, \widehat{Y}_h^*)$. Let $[\mathbf{u}_0, p_0] \in \widehat{X}_h$ be the unique solution of

$$\langle B([\mathbf{u}_0, p_0]), [\mathbf{w}, r]\rangle = \langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\mathbf{w}, r]\rangle \quad \forall[\mathbf{w}, r] \in \widetilde{Y}_h, \tag{8.16}$$

and define

$$\eta_{x_0} := \{\nu\|\mathbf{u}_0\|_{1,2;\omega_0}^2 + \|p_0\|_{0,2;\omega_0}^2\}^{1/2}. \tag{8.17}$$

Note, that problem (8.16) is equivalent to

$$\nu\int_{\omega_0} \nabla\mathbf{u}_0\nabla\mathbf{w} - \int_{\omega_0} p_0\nabla\cdot\mathbf{w} = \nu\int_{\omega_0}\nabla\mathbf{u}_h\nabla\mathbf{w} + \int_{\omega_0}(\mathbf{u}_h\cdot\nabla)\mathbf{u}_h\mathbf{w} - \int_{\omega_0} p_h\nabla\cdot\mathbf{w} - \int_{\omega_0}\pi_{0,T}\mathbf{f}\mathbf{w} \quad \forall\mathbf{w} \in M_0,$$
$$\int_{\omega_0} r\nabla\cdot\mathbf{u}_0 = \int_{\omega_0} r\nabla\cdot\mathbf{u}_h \quad \forall r \in Q_0.$$

This shows that η_{x_0} falls into the class of a posteriori error estimators originally introduced in [4, 5] for the Poisson equation.

Obviously, we have $\widetilde{Y}_h|_{\omega_0} \subset \widehat{Y}_h$. Lemma 5.1 and equation (8.8), on the other hand, immediately imply

$$\begin{aligned}
&\|\widetilde{F}_h([\mathbf{u}_h, p_h])\|_{\widehat{Y}_h^*} \\
&= \sup_{\substack{[\mathbf{v},q]\in\widehat{Y}_h \\ \|[\mathbf{v},q]\|_Y=1}} \sum_{T\in\mathcal{T}_0}\left\{\int_T\{-\nu\Delta\mathbf{u}_h + (\mathbf{u}_h\cdot\nabla)\mathbf{u}_h + \nabla p_h - \pi_{0,T}\mathbf{f}\}\mathbf{v} + \int_T q\nabla\cdot\mathbf{u}_h\right\} \\
&\qquad + \sum_{E\in\mathcal{E}_0}\int_E[\nu\partial_n\mathbf{u}_h - p_h\mathbf{n}]_E\mathbf{v} \\
&\le c\left\{\sum_{T\in\mathcal{T}_0}\eta_T^2\right\}^{1/2}.
\end{aligned}$$

Together with inequality (8.14), this proves

$$\|\widetilde{F}_h([\mathbf{u}_h, p_h])\|_{\widehat{Y}_h^*} \le c \sup_{\substack{[\mathbf{v},q]\in\widetilde{Y}_h|_{\omega_0} \\ \|[\mathbf{v},q]\|_Y=1}} \langle \widetilde{F}_h([\mathbf{u}_h, p_h]), [\mathbf{v}, q]\rangle.$$

These results and Proposition 4.3 yield the following proposition, which is a generalization of results in [4, 5, 25, 27, 28].

Proposition 8.4. *Let x_0 be an arbitrary vertex in the partition $\mathcal{T}_h$. Then there are two constants c_1, c_2, which only depend on the polynomial degree of the spaces M_h, Q_h and on the ratio h_T/ϱ_T, such that the following inequalities hold:*

$$c_1 \left\{ \sum_{T\in\mathcal{T}_0} \eta_T^2 \right\}^{1/2} \le \eta_{x_0} \le c_2 \left\{ \sum_{T\in\mathcal{T}_0} \eta_T^2 \right\}^{1/2} .$$

Here, η_T and η_{x_0} are given by equations (8.7) *and* (8.17), *respectively.*

Bibliography

1. R. A. Adams, *Sobolev spaces*, Academic Press, New York, 1975.

2. I. Babuška, *Feedback, adaptivity, and a posteriori estimates in finite elements: aims, theory, and experience*, Accuracy Estimates and Adaptive Refinements in Finite Element Computation (I. Babuška et al., eds.), Wiley, New York, 1986, pp. 3–23.

3. I. Babuška and W. Gui, *Basic principles of feedback and adaptive approaches in the finite element method*, Comput. Methods Appl. Mech. Engrg. **55** (1986), 27–42.

→ I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal. **15** (1978), 736–754.

5. → ___, *A posteriori error estimates for the finite element method*, Internat. J. Numer. Methods Engrg. **12** (1978), 1597–1615.

6. I. Babuška and R. Rodriguez, *The problem of the selection of an a-posteriori error indicator based on smoothing techniques*, Internat. J. Numer. Methods. Engrg. (to appear).

7. R. E. Bank and A. Weiser, *Some a posteriori error estimators for elliptic partial differential equations*, Math. Comp. **44** (1985), 283–301.

8. R. E. Bank and D. B. Welfert, *A posteriori error estimates for the Stokes equations: a comparison*, Comput. Methods Appl. Mech. Engrg. **87** (1990), 323–340.

9. J. Baranger and H. El Amri, *Estimateur a posteriori d'erreur pour le calcul adaptif d'écoulements quasi-Newtoniens*, RAIRO Modél. Math. Anal. Numér. **25** (1991), 31–48.

10. Ph. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, Amsterdam, 1978.

11. Ph. Clément, *Approximation by finite element functions using local regularization*, RAIRO Modél. Math. Anal. Numér. **2** (1975), 77–84.

12. M. Crouzeix and J. Rappaz, *On numerical approximation in bifurcation theory*, Research in Appl. Math., vol. 13, Masson, Paris; Springer-Verlag, Berlin, 1990.

→ R. Duran, M. A. Muschietti, and R. Rodriguez, *On the asymptotic exactness of error estimators for linear triangular elements*, Numer. Math. **59** (1991), 107–127.

14. R. Duran and R. Rodriguez, *On the asymptotic exactness of Bank-Weiser's estimator*, Numer. Math. **62** (1992), 297–303.

15. K. Eriksson, *Improved accuracy by adapted mesh-refinements in the finite element method*, Math. Comp. **44** (1985), 321–343.

16. K. Eriksson and C. Johnson, *An adaptive finite element method for linear elliptic problems*, Math. Comp. **50** (1988), 361–383.

17. V. Girault and P.-A. Raviart, *Finite element methods for Navier-Stokes equations: Theory and algorithms*, Springer Ser. Comput. Math., vol. 5, Springer, Berlin, 1986.

18. J. T. Oden, L. Demkowicz, W. Rachowicz, and T. A. Westermann, *Toward a universal $h-p$ adaptive finite element strategy, Part* 2. *A posteriori error estimation*, Comput. Methods Appl. Mech. Engrg. **77** (1989), 113–180.

19. J. Pousin and J. Rappaz, *Consistency, stability, a priori and a posteriori errors for Petrov-Galerkin methods applied to nonlinear problems* (Report), EPFL, Lausanne, 1992.

20. W. C. Rheinboldt, *On a theory of mesh-refinement processes*, SIAM J. Numer. Anal. **17** (1980), 766–778.

21. T. Strouboulis and K. A. Hague, *Recent experiences with error estimation and adaptivity* I: *Review of error estimators for scalar elliptic problems*, Comput. Methods Appl. Mech. Engrg. **97** (1992), 399–436.

22. T. Strouboulis and J. T. Oden, *A posteriori error estimation of finite element approximations in fluid mechanics*, Comput. Methods Appl. Mech. Engrg. **78** (1990), 201–242.

23. L. Tobiska and R. Verfürth, *Analysis of a streamline diffusion finite element method for the Stokes and Navier-Stokes equations* (Report), Universities Magdeburg-Zürich, 1991.

24. R. Verfürth, *Finite element approximation of incompressible Navier-Stokes equations with slip boundary conditions*, Numer. Math. **50** (1987), 697–721.

25. ——, *A posteriori error estimators for the Stokes equations*, Numer. Math. **55** (1989), 309–325.

26. ——, *Finite element approximation of incompressible Navier-Stokes equations with slip boundary conditions*. II, Numer. Math. **59** (1991), 615–636.

27. ——, *A posteriori error estimators and adaptive mesh-refinement techniques for the Navier-Stokes equations*, Incompressible CFD-Trends and Advances (M. D. Gunzburger and R. A. Nicolaides, eds.), Cambridge Univ Press, Cambridge, 1993, pp. 447–477.

28. ——, *A posteriori error estimation and adaptive mesh-refinement techniques*, J. Comput. Appl. Math. (to appear).

29. J. Z. Zhu and O. C. Zienkiewicz, *Adaptive techniques in the finite element method*, Comm. Appl. Numer. Methods **4** (1988), 197–204.

30. O. C. Zienkiewicz and J. Z. Zhu, *A simple error estimator and adaptive procedure for practical engineering analysis*, Internat. J. Numer. Methods Engrg. **24** (1987), 337–357.

Fakultät für Mathematik, Ruhr-universität Bochum, D-44780 Bochum, Germany