Mathematical Methods in the Applied Sciences, Vol. 17, 487–508 (1994)
MOS subject classification: 35 A 15, 65 G 99, 49 A 55, 35 J 65

A posteriori error analysis for linearization of nonlinear elliptic problems and their discretizations

Weimin Han

Department of Mathematics, University of Iowa, Iowa City, IA 52242, USA

Communicated by R. P. Gilbert

The paper is devoted to *a posteriori* quantitative analysis for errors caused by linearization of non-linear elliptic boundary value problems and their finite element realizations. We employ duality theory in convex analysis to derive computable bounds on the difference between the solution of a non-linear problem and the solution of the linearized problem, by using the solution of the linearized problem only. We also derive computable bounds on differences between finite element solutions of the nonlinear problem and finite element solutions of the linearized problem, by using finite element solutions of the linearized problem only. Numerical experiments show that our *a posteriori* error bounds are efficient.

1. Introduction

The aim of scientific computation is to reliably describe and predict physical phenomena of interest. The reliability of a numerical solution of a physical or engineering problem depends on mathematical idealization of the physical problem and numerical treatment of the idealized mathematical problem. An illuminating description of the basic flow chart of numerical analysis of a physical problem is found in [2]. To analyse a physical problem, a very first stage is to establish a basic mathematical formulation for the problem. It is a highly idealized assumption that we can have a mathematical problem which exactly describes the physical problem. The available data, which usually come from experiments, for the basic mathematical formulation cannot be obtained as accurately as one wishes. As a consequence, we solve a simplified, or idealized mathematical problem, instead. It is strongly desirable that we should be able to analyse the reliability and the error of the solution resulted from the idealization. This procedure forms the second stage. The next stage is to employ certain numerical method to solve the idealized mathematical problem, and analyse the reliability and the error of the numerical solution. Finally, if both the mathematical idealization and the numerical solution of the idealized problem are reliable, various information for the real problem is drawn based on the numerical solution of the idealized problem.

There is a large amount of literature on numerical methods and their error analysis. Relatively few work has been done for reliability analysis of mathematical idealizations, especially most desirably, certain good, practically useful quantitative assess-

CCC 0170–4214/94/070487–22 *Received 27 January 1993*

ments of the quality of solutions of idealized problems. Such quantitative assessments should be (hopefully) available once we have computed the solutions of idealized problems. One should not assume the knowledge of the solutions of basic mathematical problems for either exact decriptions of basic mathematical problems are usually not available in practice or, it is often too expensive to solve the basic mathematical problems.

This paper is one in a series devoted to systematically analyse effects of mathematical idealizations on solutions, and to provide *a posteriori* quantitative error estimates for the solutions of certain idealized problems. We employ the duality theory in convex analysis for *a posteriori* error analysis of solutions of idealized mathematical problems. The method used can be viewed as a systematic development of the conventional two-sided energy estimate technique. For some references on two-sided energy estimate technique, the reader is referred to [3, 6–8, 10, 14] and references therein. There are different approaches to establish a dual variational principle. We use the duality theory in convex analysis (cf. [4]) for this purpose. For some other approaches, see [1, 12–14].

In this paper, we consider one kind of constitutive law idealization: linearization. Detailed quantitative error analysis is given for linearizations of certain non-linear elliptic problems, whose linearizations are boundary value problems of Poisson's equations. An example of a non-linear problem considered in this paper is the following:

$$\begin{aligned} -\operatorname{div}(a(|\nabla u|^2)\nabla u) &= f \quad \text{in } \Omega, \\ u &= g \quad \text{on } \partial\Omega. \end{aligned} \tag{1.1}$$

When $\Omega \in \mathbb{R}^2$ is an open bounded domain, for suitable data, a, f and g, problem (1.1) describes a non-linear torsion problem (cf. [11]). In particular, when the coefficient function $a(\xi) \equiv 1$, the problem reduces to

$$\begin{aligned} -\Delta u_0 &= f \quad \text{in } \Omega, \\ u_0 &= g \quad \text{on } \partial\Omega, \end{aligned} \tag{1.2}$$

which is a linear torsion problem. Owing to the difficulty associated with finding the material property function $a(\xi)$, in most elasticity theory books, it is taken for granted that the torsion problem of a real material has been described accurately enough by the linearized problem (1.2), since the coefficient function $a(\xi)$ is usually close to some constant (taken as 1 here) for many commonly seen materials. In this paper, we give *a posteriori* quantitative error estimates for the effect of linearization on solutions of elliptic problems, using solutions of the linearized problems only. In particular, we provide error bounds on $u - u_0$ for the solutions of (1.1) and (1.2) in terms of the solution u_0 of the linearized problem (1.1).

In practice, numerical methods are used to solve problems (1.1) and (1.2). Like the continuous problems, it is usually not easy to solve problem (1.1) numerically, and we may first solve the simpler problem (1.2) numerically and check if the numerical solution of (1.2) is close to, and can be used as a good approximation of the solution of (1.1). For finite element solutions of problems (1.1) and (1.2), we will provide a discrete version of *a posteriori* estimates for the error caused by the linearization, i.e. some computable error bounds for the difference between a finite element solution of problem (1.1) and a finite element solution of problem (1.2), using the finite element solution of problem (1.2) only.

The organization of the paper is as follows. In section 2, we review some standard results from duality theory in convex analysis, and provide a general framework for *a posteriori* error estimates. In section 3, we derive a quantitative bound for the difference between the solution of a nonlinear problem and that of a related linearized problem. The quantitative error bound is given in terms of the solution of the linearized problem and an auxiliary function subjected to some constraint. We discuss the problem on selecting suitable auxiliary functions in section 4. In section 5, we give a discrete version of *a posteriori* error estimates for finite element solutions of problems (1.1) and (1.2). Numerical experiments are presented in section 6 showing that our error bounds are efficient. Finally, some further discusions are given concerning *a posteriori* error analysis for effects of mathematical idealizations on solutions.

2. Duality theory, *a posteriori* error estimate

The results of the first two subsections can be found in [4]. We include them here for readers' convenience.

2.1. *Convexity, conjugate functions*

Let V be a real normed vector space, $F: V \to \bar{\mathbb{R}}$ be a mapping which can take the values $+\infty$ and $-\infty$. F is said to be *convex* if

$$F(\lambda u + (1-\lambda)v) \leqslant \lambda F(u) + (1-\lambda)F(v), \quad \forall u, v \in V, \ \forall \lambda \in [0, 1], \tag{2.1}$$

whenever the right-hand side is defined. F is said to be strictly convex if the strict inequality in (2.1) holds for any $u \neq v$, $\lambda \in (0, 1)$. F is said to be *proper* if $F \not\equiv +\infty$, and $F(v) \neq -\infty$, $\forall v \in V$. F is said to be *lower semi-continuous* on V if $\forall a \in \mathbb{R}$, $\{v \in V | F(v) \leqslant a\}$ is closed.

Let V^* be the dual space of V with duality pairing $\langle \cdot, \cdot \rangle$. The conjugate function $F^*: V^* \to \bar{\mathbb{R}}$ of F is defined by

$$F^*(v^*) = \sup_{v \in V} [\langle v, v^* \rangle - F(v)]. \tag{2.2}$$

Often, we shall have to calculate the conjugate function for a function defined by an integral of the form

$$G(v) = \int_\Omega g(x, v(x))\, dx.$$

Before stating a theorem on how to calculate its conjugate function, we introduce the following notion.

Definition 2.1. *Let Ω be an open set of $\mathbb{R}^n$, $g: \Omega \times \mathbb{R}^l \to \mathbb{R}$, g is said to be a Carathéodory function if*

$$\forall \xi \in \mathbb{R}^l, \quad x \mapsto g(x, \xi) \text{ is a measurable function,}$$

$$\text{for a.e. } x \in \Omega, \quad \xi \mapsto g(x, \xi) \text{ is a continuous function.}$$

Let $m_i \in (1, \infty)$, $i = 1, \ldots, l$. For each i, denote

$$L^{m_i}(\Omega) = \left\{ v \text{ measurable} \;\middle|\; \int_\Omega |v|^{m_i}\,\mathrm{d}x < \infty \right\}.$$

We have the following theorem.

Theorem 2.2. *Let $g : \Omega \times \mathbb{R}^l \to \mathbb{R}$ be a Carathéodory function. For any $v \in V = \prod_{i=1}^{l} L^{m_i}(\Omega)$, define*

$$G(v) = \int_\Omega g(x, v(x))\,\mathrm{d}x.$$

Then for the conjugate function of G,

$$G^*(v^*) = \int_\Omega g^*(x, v^*(x))\,\mathrm{d}x, \quad \forall v^* \in V^*,$$

where

$$g^*(x, y) = \sup_{\xi \in \mathbb{R}^l} [y \cdot \xi - g(x, \xi)].$$

In our calculation later, we will use Gâteaux differentials. Recall that F is said to be *Gâteaux differentiable* at u if $\exists u^* \in V^*$, such that

$$\lim_{\lambda \to 0} \frac{F(u + \lambda v) - F(u)}{\lambda} = \langle v, u^* \rangle, \quad \forall v \in V,$$

u^* is called the Gâteaux differential of F at u, and is denoted by $F'(u)$.

2.2. *Duality theory*

Let V, Q be two normed spaces, V^*, Q^* their dual spaces. Assume there exists a linear continuous operator $\Lambda \in \mathscr{L}(V, Q)$, with transpose $\Lambda^* \in \mathscr{L}(Q^*, V^*)$. Let J be a function of $V \times Q$ into $\bar{\mathbb{R}}$. Consider the minimization problem:

$$\inf_{v \in V} J(v, \Lambda v). \tag{2.3}$$

Define its dual problem by

$$\sup_{q^* \in Q^*} [-J^*(\Lambda^* q^*, -q^*)], \tag{2.4}$$

where J^* is the conjugate function of J:

$$J^*(v^*, q^*) = \sup_{v \in V, q \in Q} [\langle v, v^* \rangle + \langle q, q^* \rangle - J(v, q)]. \tag{2.5}$$

For the relation between problems (2.3) and (2.4), we have the following duality theorem.

Theorem 2.3. *Assume*

(1) *V is a reflexive Banach space, Q a normed space,*

(2) *$J : V \times Q \to \bar{\mathbb{R}}$ is a proper, lower semi-continuous, convex function,*

(3) *$\exists u_0 \in V$ such that $J(u_0, \Lambda u_0) < \infty$ and $q \mapsto J(u_0, q)$ is continuous at Λu_0,*
(4) *$J(v, \Lambda v) \to +\infty$, as $\|v\| \to \infty$, $v \in V$.*

Then problem (2.3) has a solution $u \in V$, problem (2.4) has a solution $p^* \in Q^*$, and

$$J(u, \Lambda u) = -J^*(\Lambda^* p^*, -p^*). \tag{2.6}$$

Furthermore, if J is strictly convex, then a solution u of problem (2.3) is unique.

If the assumptions are weakened, then a weaker form of the theorem holds.

Theorem 2.4. *Assume*

(1) *V, Q are normed spaces,*
(2) *$J: V \times Q \to \bar{\mathbb{R}}$ is convex,*
(3) *$\exists u_0 \in V$ such that $J(u_0, \Lambda u_0) < \infty$ and $q \mapsto J(u_0, q)$ is continuous at Λu_0,*
(4) *$\inf_{v \in V} J(v, \Lambda v)$ is finite.*

Then problem (2.4) has a solution $p^* \in Q^*$, and

$$\inf_{v \in V} J(v, \Lambda v) = -J^*(\Lambda^* p^*, -p^*). \tag{2.7}$$

Furthermore, if J is strictly convex, then a solution u (if it exists) of problem (2.3) is unique.

We will often encounter the situation where the function J is of a separated form, i.e.

$$J(v, q) = F(v) + G(q). \tag{2.8}$$

It is easily calculated that the conjugate function of J is

$$J^*(v^*, q^*) = F^*(v^*) + G^*(q^*), \tag{2.9}$$

where F^*, G^* are the conjugate functions of F, G. Specializing Theorem 2.3 in this case, we have the following theorem.

Theorem 2.5. *Assume*

(1) *V is a reflexive Banach space, Q a normed space,*
(2) *$F: V \to \bar{\mathbb{R}}$, $G: Q \to \bar{\mathbb{R}}$ are proper, lower semi-continuous, convex functions,*
(3) *$\exists u_0 \in V$ such that $F(u_0) < \infty$, $G(\Lambda u_0) < \infty$, $q \mapsto G(q)$ is continuous at Λu_0,*
(4) *$F(v) + G(\Lambda v) \to +\infty$ as $\|v\| \to \infty$, $v \in V$.*

Then the conclusions of Theorem 2.3 hold.

2.3. *A general framework for a posteriori error estimates*

Let $u \in V$ be a solution of the minimization problem (2.3). Assume J is Gâteaux differentiable at u. Let $v \in V$ be any element with $J(v, \Lambda v) < \infty$. We set

$$D(u, v) = J(v, \Lambda v) - J(u, \Lambda u) - \langle J'(u, \Lambda u), (v - u, \Lambda v - \Lambda u) \rangle, \tag{2.10}$$

where $\langle \cdot, \cdot \rangle$ is the duality pairing between $(V \times Q)^*$ and $V \times Q$.

Theorem 2.6. *Make the same assumptions as in Theorem 2.3 (or in Theorem 2.4) together with the existence of the solution of problem (2.3). If J is Gâteaux differentiable at u, then*

$$D(u, v) \leqslant J(v, \Lambda v) + J^*(\Lambda^* q^*, -q^*), \quad \forall q^* \in Q^*. \tag{2.11}$$

Proof. From the inequality

$$J(u+\lambda(v-u),\ \Lambda u+\lambda(\Lambda v-\Lambda u))\geqslant J(u,\Lambda u),\quad \forall\lambda\in(0,1),$$

we claim that

$$\langle J'(u,\Lambda u),(v-u,\Lambda v-\Lambda u)\rangle\geqslant 0. \tag{2.12}$$

Thus, applying Theorem 2.3 or Theorem 2.4 (with p^* a solution of the dual problem (2.4)),

$$\begin{aligned} D(u,v) &\leqslant J(v,\Lambda v)-J(u,\Lambda u)\\ &= J(v,\Lambda v)-[-J^*(\Lambda^*p^*,\ -p^*)]\\ &\leqslant J(v,\Lambda v)+J^*(\Lambda^*q^*,\ -q^*),\quad \forall q^*\in Q^*. \end{aligned}$$

□

Remark 2.7. The inequality (2.12) holds even if V is assumed to be a convex subset of a normed space. When V is a normed space, (2.12) is actually an equality, as can easily be seen by replacing v by $2u-v$ in (2.12). Hence, if V is a normed space, we have the equality

$$D(u,v)=J(v,\Lambda v)-J(u,\Lambda u).$$

Armed with Theorem 2.6, the procedure of deriving an estimate for the difference between u and v is decomposed into two steps.

Step 1. Find a suitable lower bound for $D(u,v)$ which measures the difference between u and v. Usually, this lower bound will be the error in energy, or some quantity depending on $\|v-u\|$.

Step 2. Take an appropriate q^* so that the estimate (2.11) is as accurate as possible. If q^* is chosen to be a solution p^* of the dual problem, then the right-hand side of (2.11) attains its minimum. However, usually it is not easy to find p^*. So it is desirable to have a strategy on determining a q^* that is easy to get and that produces a good bound for the right-hand side of (2.11).

To use Theorem 2.6 for an error estimate for the linearization of the problem (1.1), we will take u to be the solution of (1.1), $v=u_0$ the solution of the linearized problem (1.2). We will construct suitable auxiliary functions q^*, based on information from the solution u_0 of the linearized problem, to produce good estimates for the error $u-u_0$.

3. *A posteriori* error estimates for linearization (continuous problems)

In this section, we use the results of last section to derive *a posteriori* error estimates for linearization of non-linear problems of the form (1.1). We state our results for two-dimensional problems only. It is straightforward to give similar results for any finite-dimensional problems.

3.1. *The non-linear problem and its linearization*

Let $\Omega\subset\mathbb{R}^2$ be an open bounded domain. Let $\partial\Omega=\bar\Gamma_1\cup\bar\Gamma_2$ with $\Gamma_1\cap\Gamma_2=\emptyset$. We assume

$$\text{meas}\,(\Gamma_1)\neq 0.$$

Let

$$f \in L^2(\Omega), \qquad g_1 \in H^1(\Omega), \qquad g_2 \in L^2(\Gamma_2).$$

We will use the same notation g_1 to denote the trace of $g_1 \in H^1(\Omega)$ on the boundary.

To have a class of problems for which we can make specific computations, we introduce the following scalar function as the coefficient function:

$$a(\xi) = \begin{cases} 1 & \text{if } \xi \leqslant \xi_0, \\ \beta + (1-\beta)\sqrt{(\xi_0/\xi)} & \text{if } \xi > \xi_0 \end{cases} \tag{3.1}$$

for some $\xi_0 > 0$, $\beta \in (0, 1)$.

We take

$$V = H^1(\Omega), \qquad V^* = (H^1(\Omega))^*;$$

$$Q = Q^* = (L^2(\Omega))^2 \times L^2(\Gamma_2);$$

$$\Lambda v = (\nabla v, v|_{\Gamma_2}).$$

Denote

$$H^1_{\Gamma_1}(\Omega) = \{v \in H^1(\Omega) | v = 0 \text{ on } \Gamma_1\},$$

$$H^1_{\Gamma_1, g_1}(\Omega) = \{v \in H^1(\Omega) | v - g_1 \in H^1_{\Gamma_1}(\Omega)\}.$$

We define the energy function on V:

$$J(v, \Lambda v) = F(v) + G(\Lambda v), \tag{3.2}$$

where

$$F(v) = \begin{cases} -\int_\Omega fv \, \mathrm{d}x & \text{if } v \in H^1_{\Gamma_1, g_1}(\Omega), \\ +\infty & \text{otherwise;} \end{cases} \tag{3.3}$$

$$G(\Lambda v) = \int_\Omega \int_0^{|\nabla v|^2} \frac{1}{2} a(\xi) \, \mathrm{d}\xi \, \mathrm{d}x - \int_{\Gamma_2} g_2 v \, \mathrm{d}s. \tag{3.4}$$

By Theorem 2.3, we have a unique solution u to the minimization problem:

$$\inf_{v \in V} J(v, \Lambda v). \tag{3.5}$$

The solution u is easily seen to be the weak solution of the non-linear elliptic boundary value problem

$$\begin{aligned} -\operatorname{div}(a(|\nabla u|^2)\nabla u) &= f \quad \text{in } \Omega. \\ u &= g_1 \quad \text{on } \Gamma_1, \\ a(|\nabla u|^2)\frac{\partial u}{\partial n} &= g_2 \quad \text{on } \Gamma_2. \end{aligned} \tag{3.6}$$

We call

$$\Omega_1(u) = \{x \in \Omega \,|\, |\nabla u(x)|^2 \leqslant \xi_0\}$$

the linear region of the non-linear solution u, and

$$\Omega_n(u) = \{x \in \Omega \,|\, |\nabla u(x)|^2 > \xi_0\}$$

the non-linear region of u. Since the solution depends continuously on the input data, when f, g_1 and g_2 are small, the non-linear region $\Omega_n(u)$ is small. Note that on $\Omega_l(u)$, $a(|\nabla u|^2) = 1$. Hence, when f, g_1 and g_2 are small, we expect that the solution u_0 of the following linear problem is a good approximation of u:

$$\begin{aligned} -\Delta u_0 &= f \quad \text{in } \Omega, \\ u_0 &= g_1 \quad \text{on } \Gamma_1, \\ \frac{\partial u_0}{\partial n} &= g_2 \quad \text{on } \Gamma_2. \end{aligned} \tag{3.7}$$

However, if Ω is a plane corner domain, e.g. if its boundary $\partial\Omega$ is smooth everywhere except at a corner O (taken as the origin of the co-ordinate system we are working on) with an internal angle $\omega > \pi$, then no matter how smooth f, g_1 and g_2 are, the solution u_0 of the linearized problem (3.7) is, in general, singular at O; for detail, see [5]. Let $\alpha = \pi/\omega$. When on both sides of the part of the boundary around O, either a Dirichlet or a Neumann boundary condition is specified, we have

$$|\nabla u_0(x)| \sim O(r^{\alpha-1}) \quad \text{for } r = |x| \text{ close to } 0.$$

When on one side a Dirichlet condition is specified, while on the other side a Neumann condition is specified, we have

$$|\nabla u_0(x)| \sim O(r^{\alpha/2-1}) \quad \text{for } r = |x| \text{ close to } 0.$$

Since $\alpha < 1$, $|\nabla u_0(x)| \to \infty$ as $x \to O$. Hence, it is doubtful whether the linearized problem (3.7) is in any sense close to the non-linear problem (3.6). In the next subsection, we derive an estimate for the difference between u and u_0.

Remark 3.1. Why the coefficient function $a(\xi)$ is chosen in the form (3.1)? The stress–strain relation corresponding to the nonlinear problem (3.6) is $\sigma = a(|\varepsilon|^2)\varepsilon$. For scalars σ and ε, we require a stress–strain relation as illustrated in Fig. 1, where we

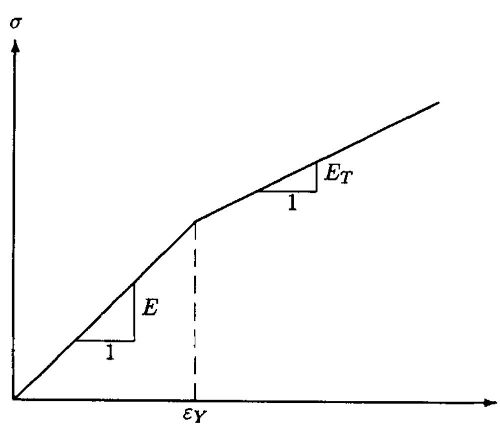

Fig. 1. Stress–strain relation (ε_Y = yield strain)

omit the part of the graph for negative σ and ε, which can be obtained by reflecting the graph for positive σ and ε with respect to the origin. Hence, the coefficient $a(\xi)$ is of the form (3.1) with $\beta = E_T/E$, $\xi_0 = \varepsilon_Y^2$.

3.2. *Quantitative error estimate*

For $q \in Q$, we write $q = (q_1, q_2)$ with $q_1 \in (L^2(\Omega))^2$ and $q_2 \in L^2(\Gamma_2)$. A similar notation is used for $q^* \in Q^*$. We now compute the conjugate functions.

$$\begin{aligned}
F^*(\Lambda^* q^*) &= \sup_{v \in V} \{\langle \Lambda v, q^* \rangle - F(v)\} \\
&= \sup_{v \in H^1_{\Gamma_1, g_1}(\Omega)} \left\{ \int_\Omega (q_1^* \nabla v + f v)\,\mathrm{d}x + \int_{\Gamma_2} q_2^* v\,\mathrm{d}s \right\} \\
&= \int_\Omega (q_1^* \nabla g_1 + f g_1)\,\mathrm{d}x + \int_{\Gamma_2} q_2^* g_1\,\mathrm{d}s \\
&\quad + \sup_{v \in H^1_0(\Omega)} \left\{ \int_\Omega (q_1^* \nabla v + f v)\,\mathrm{d}x + \int_{\Gamma_2} q_2^* v\,\mathrm{d}s \right\} \\
&\quad = \begin{cases} \displaystyle\int_\Omega (q_1^* \nabla g_1 + f g_1)\,\mathrm{d}x + \int_{\Gamma_2} q_2^* g_1\,\mathrm{d}s & \text{if } \operatorname{div} q_1^* = f \text{ in } \Omega, \\ +\infty \text{ otherwise}; \ -q_1^* n = q_2 \text{ on } \Gamma_2, \end{cases}
\end{aligned}$$

$$\begin{aligned}
G^*(-q^*) &= \sup_{q \in Q} \{\langle q, -q^* \rangle - G(q)\} \\
&= \sup_{q \in Q} \left\{ \int_\Omega -q_1^* q_1\,\mathrm{d}x - \int_\Omega \int_0^{|q_1|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x \right. \\
&\quad \left. + \int_{\Gamma_2} (-q_2^* q_2 + g_2 q_2)\,\mathrm{d}s \right\} \\
&= \int_\Omega \sup_{q_1 \in \mathbb{R}^2} \left\{ -q_1^*(x) q_1 - \int_0^{|q_1|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi \right\} \mathrm{d}x + \int_{\Gamma_2} \sup_{q_2 \in \mathbb{R}} (g_2 - q_2^*) q_2\,\mathrm{d}s \\
&= \int_\Omega \sup_{t \geq 0} \left\{ -q_1^*(x)(-t) \frac{q_1^*(x)}{|q_1^*(x)|} - \int_0^{t^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi \right\} \mathrm{d}x \\
&\quad + \int_{\Gamma_2} \sup_{q_2 \in \mathbb{R}} (g_2 - q_2^*) q_2\,\mathrm{d}s \\
&= \int_\Omega \sup_{t \geq 0} \left\{ |q_1^*(x)| t - \int_0^{t^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi \right\} \mathrm{d}x \\
&\quad + \int_{\Gamma_2} \sup_{q_2 \in \mathbb{R}} (g_2 - q_2^*) q_2\,\mathrm{d}s \\
&= \int_\Omega \left\{ |q_1^*| b(|q_1^*|) - \int_0^{b(|q_1^*|)^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi \right\} \mathrm{d}x \quad \text{if } q_2^* = g_2 \text{ on } \Gamma_2,
\end{aligned}$$

where $t = b(|q_1^*|)$ is the solution of the equation:

$$|q_1^*| = a(t^2)t,$$

thus,

$$b(\eta) = \begin{cases} \eta & \text{if } \eta \leqslant \sqrt{\xi_0}, \\ (\eta - (1-\beta)\sqrt{\xi_0})/\beta & \text{if } \eta > \sqrt{\xi_0}. \end{cases} \tag{3.8}$$

In computing $G^*(-q^*)$, we used Theorem 2.2, also we assumed $q_1^* \neq 0$. Obviously, the expression for $G^*(-q^*)$ holds for $q_1^* = 0$, too.

Thus,

$$\begin{aligned} J^*(\Lambda^* q^*, -q^*) &= F^*(\Lambda^* q^*) + G^*(-q^*) \\ &= \int_\Omega \left(q_1^* \nabla g_1 + f g_1 + |q_1^*| b(|q_1^*|) - \int_0^{b(|q_1^*|)^2} \tfrac{1}{2} a(\xi)\,\mathrm{d}\xi \right) \mathrm{d}x \\ &\quad + \int_{\Gamma_2} q_2^* g_1 \,\mathrm{d}s \quad \text{if } \operatorname{div} q_1^* = f \text{ in } \Omega, \quad -q_1^* n = q_2^* = g_2 \text{ on } \Gamma_2, \\ &\quad + \infty \quad \text{otherwise.} \end{aligned}$$

Using Theorem 2.6, we then have the following estimate for the difference in energy:

$$\begin{aligned} J(u_0, \Lambda u_0) - J(u, \Lambda u) &\leqslant \int_\Omega \int_0^{|\nabla u_0|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x - \int_\Omega f u_0 \,\mathrm{d}x - \int_{\Gamma_2} g_2 v \,\mathrm{d}s \\ &\quad + \int_\Omega \left(q_1^* \nabla g_1 + f g_1 + |q_1^*| b(|q_1^*|) \right. \\ &\quad \left. - \int_0^{b(|q_1^*|)^2} \frac{1}{2} a(\xi)\mathrm{d}\xi \right) \mathrm{d}x + \int_{\Gamma_2} g_2 g_1 \,\mathrm{d}s \\ &= \int_\Omega \int_{b(|q_1^*|)^2}^{|\nabla u_0|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x + \int_\Omega (q_1^* \nabla g_1 + |q_1^*| b(|q_1^*|))\mathrm{d}x \\ &\quad - \int_\Omega f(u_0 - g_1)\,\mathrm{d}x - \int_{\Gamma_2} g_2(u_0 - g_1)\,\mathrm{d}s \\ &= \int_\Omega \int_{b(|q_1^*|)^2}^{|\nabla u_0|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x + \int_\Omega (q_1^* \nabla g_1 + |q_1^*| b(|q_1^*|))\,\mathrm{d}x \\ &\quad - \int_\Omega \nabla u_0 \nabla(u_0 - g_1)\,\mathrm{d}x \\ &= \int_\Omega \int_{b(|q_1^*|)^2}^{|\nabla u_0|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x + \int_\Omega \Big((q_1^* + \nabla u_0) \nabla g_1 \\ &\quad + |q_1^*| b(|q_1^*|) - |\nabla u_0|^2 \Big) \mathrm{d}x, \\ &\quad \forall q_1^* \in (L^2(\Omega))^2, \ \operatorname{div} q_1^* = f \text{ in } \Omega, \quad -q_1^* n = g_2 \text{ on } \Gamma_2. \end{aligned}$$

Since in the above estimate, the second component q_2^* of the auxiliary function q^* does not play a role, we will use the notation q^* for the first component q_1^*. Hence, the above estimate is rewritten as

$$J(u_0, \Lambda u_0) - J(u, \Lambda u)$$
$$\leqslant \int_\Omega \int_{b(|q^*|)^2}^{|\nabla u_0|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x + \int_\Omega \Big((q^* + \nabla u_0)\nabla g_1 + |q^*| b(|q^*|) - |\nabla u_0|^2 \Big)\mathrm{d}x, \tag{3.9}$$
$$\forall q^* \in (L^2(\Omega))^2,\ \operatorname{div} q^* = f \text{ in } \Omega,\quad -q^* n = g_2 \text{ on } \Gamma_2.$$

Now we consider the quantity $D(u, u_0)$. Denote

$$\Omega_l(u_0) = \{x \in \Omega \mid |\nabla u_0(x)| \leqslant \sqrt{\xi_0}\},$$
$$\Omega_n(u_0) = \{x \in \Omega \mid |\nabla u_0(x)| > \sqrt{\xi_0}\},$$
$$\Omega_l(u, u_0) = \{x \in \Omega \mid |\nabla u(x)|, |\nabla u_0(x)| \leqslant \sqrt{\xi_0}\},$$
$$\Omega_n(u, u_0) = \{x \in \Omega \mid |\nabla u(x)|, |\nabla u_0(x)| > \sqrt{\xi_0}\},$$

the linear and non-linear regions of u_0 and of both u and u_0. We have

$$D(u, u_0) = J(u_0, \Lambda u_0) - J(u, \Lambda u) - \langle J'(u, \Lambda u), (u_0 - u, \Lambda u_0 - \Lambda u)\rangle$$
$$= G(\Lambda u_0) - G(\Lambda u) - \langle G'(\Lambda u), \Lambda(u_0 - u)\rangle$$
$$= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} G(\Lambda u + t\Lambda(u_0 - u))\,\mathrm{d}t - \langle G'(\Lambda u), \Lambda(u_0 - u)\rangle$$
$$= \int_0^1 \langle G'(\Lambda u + t\Lambda(u_0 - u)) - G'(\Lambda u), \Lambda(u_0 - u)\rangle\,\mathrm{d}t$$
$$= \int_0^1 \int_\Omega (a(|\nabla u + t\nabla(u_0 - u)|^2)(\nabla u + t\nabla(u_0 - u))$$
$$- a(|\nabla u|^2)\nabla u)\nabla(u_0 - u)\,\mathrm{d}x\,\mathrm{d}t$$
$$= \int_0^1 \int_\Omega \int_0^1 \frac{\mathrm{d}}{\mathrm{d}\tau}[a(|\nabla u + t\tau\nabla(u_0 - u)|^2)(\nabla u + t\tau\nabla(u_0 - u))]$$
$$\times \nabla(u_0 - u)\,\mathrm{d}\tau\,\mathrm{d}x\,\mathrm{d}t.$$

If $|\nabla u + t\tau\nabla(u_0 - u)| \leqslant \sqrt{\xi_0}$, then the integrand is $t|\nabla(u_0 - u)|^2$. When $|\nabla u + t\tau\nabla(u_0 - u)| > \sqrt{\xi_0}$, using the definition of a, we see that the integrand is not less than $\beta t|\nabla(u_0 - u)|^2$. Now, on $\Omega_l(u, u_0)$, $|\nabla u + t\tau\nabla(u_0 - u)| \leqslant \sqrt{\xi_0}$, thus,

$$D(u, u_0) = J(u_0, \Lambda u_0) - J(u, \Lambda u) - \langle J'(u, \Lambda u), (u_0 - u, \Lambda u_0 - \Lambda u)\rangle$$
$$\geqslant \frac{1}{2}\|\nabla(u - u_0)\|^2_{L^2(\Omega_l(u,u_0))} + \frac{\beta}{2}\|\nabla(u - u_0)\|^2_{L^2(\Omega\setminus\Omega_l(u,u_0))}. \tag{3.10}$$

Hence, by (3.9), we obtain the following estimate:

$$\|\nabla(u - u_0)\|^2_{L^2(\Omega_l(u,u_0))} + \beta\|\nabla(u - u_0)\|^2_{L^2(\Omega\setminus\Omega_l(u,u_0))}$$
$$\leqslant \int_\Omega \int_{b(|q^*|)^2}^{|\nabla u_0|^2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x + 2\int_\Omega ((q^* + \nabla u_0)\nabla g_1 + |q^*| b(|q^*|) - |\nabla u_0|^2)\,\mathrm{d}x, \tag{3.11}$$
$$\forall q^* \in (L^2(\Omega))^2,\ \operatorname{div} q^* = f \text{ in } \Omega,\quad -q^* n = g_2 \text{ on } \Gamma_2.$$

4. Selection of the auxiliary function q^*

We state two kinds of selections for q^* in the estimates (3.9) and (3.11).

Selection 1. We simply take

$$q^* = -\nabla u_0. \tag{4.1}$$

It is an admissible function, i.e. it satisfies: $\operatorname{div} q^* = f$ in Ω and $-q^* n = g_2$ on Γ_2. Then (3.9) says

$$\begin{aligned}
J(u_0, \Lambda u_0) - J(u, \Lambda u) &\leqslant \int_\Omega \int_{b(|\nabla u_0|)^2}^{|\nabla u_0|^2} \frac{1}{2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x + \int_\Omega (|\nabla u_0|\, b(|\nabla u_0|) - |\nabla u_0|^2)\,\mathrm{d}x \\
&= \int_{\Omega_n(u_0)} \left\{ \int_{b(|\nabla u_0|)^2}^{|\nabla u_0|^2} \frac{1}{2}(\beta + (1-\beta)\sqrt{(\xi_0/\xi)})\mathrm{d}\xi \right. \\
&\qquad \left. + |\nabla u_0|\, b(|\nabla u_0|) - |\nabla u_0|^2 \right\} \mathrm{d}x \\
&= \int_{\Omega_n(u_0)} \left\{ \left(\frac{\beta}{2} - 1\right)|\nabla u_0|^2 + (1-\beta)\sqrt{\xi_0}\,|\nabla u_0| \right. \\
&\qquad \left. - \frac{\beta}{2} b(|\nabla u_0|)^2 + (|\nabla u_0 - (1-\beta)\sqrt{\xi_0})\, b(|\nabla u_0|) \right\} \mathrm{d}x \\
&= \int_{\Omega_n(u_0)} \left\{ \left(\frac{\beta}{2} - 1\right)|\nabla u_0|^2 + (1-\beta)\sqrt{\xi_0}\,|\nabla u_0| \right. \\
&\qquad - \frac{\beta}{2}\frac{1}{\beta^2}(|\nabla u_0| - (1-\beta)\sqrt{\xi_0})^2 \\
&\qquad \left. + (|\nabla u_0| - (1-\beta)\sqrt{\xi_0})\frac{1}{\beta}(|\nabla u_0| - (1-\beta)\sqrt{\xi_0}) \right\} \mathrm{d}x \\
&= \frac{(1-\beta)^2}{2\beta} \int_{\Omega_n(u_0)} (|\nabla u_0| - \sqrt{\xi_0})^2\,\mathrm{d}x.
\end{aligned} \tag{4.2}$$

And, (3.11) gives

$$\begin{aligned}
&\|\nabla(u - u_0)\|^2_{L^2(\Omega_l(u, u_0))} + \beta \|\nabla(u - u_0)\|^2_{L^2(\Omega \backslash \Omega_l(u, u_0))} \\
&\leqslant \frac{(1-\beta)^2}{\beta} \int_{\Omega_n(u_0)} (|\nabla u_0| - \sqrt{\xi_0})^2\,\mathrm{d}x.
\end{aligned} \tag{4.3}$$

For β not close to 0, numerical experiments below will show that the selection (4.1) provides a good error bound. However, if β is close to 0, then the factor $1/\beta$ in (4.2) and (4.3) leads to useless error bound. To overcome this difficulty, we introduce another selection.

Selection 2. We describe this selection for a problem on a corner domain as shown in Fig. 2.

First consider the case when the boundary condition around the corner is of Dirichlet type. We assume there is an $r_0 > 0$ such that in $\Omega_0 = \Omega \cap \{r > r_0\}$, $|\nabla u_0| \leqslant \sqrt{\xi_0}/\sqrt{[(\omega/2)^2 + 1]}$, where ω is the internal angle of the corner. Denote

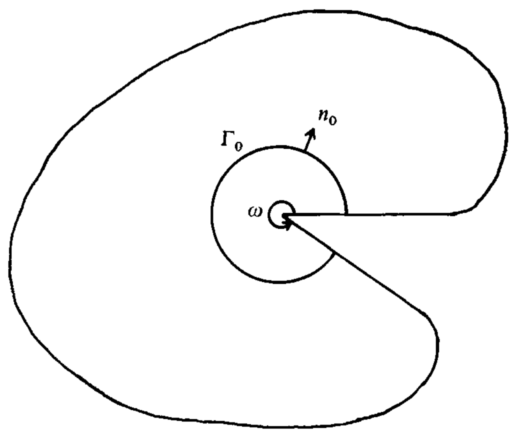

Fig. 2. A typical plane corner domain

$\Gamma_0 = \Omega \cap \{r = r_0\}$, n_0 the unit normal vector on Γ_0 in the direction of being away from O. We further assume

$$f(x) = 0, \qquad x \in \Omega \backslash \Omega_0. \tag{4.4}$$

We define a function in $\Omega \backslash \Omega_0$:

$$V(r, \theta) = r \int_{\omega/2}^{\theta} \left. \frac{\partial u_0}{\partial n_0} \right|_{\Gamma_0} d\theta, \qquad 0 \leqslant \theta \leqslant \omega,\ 0 < r \leqslant r_0. \tag{4.5}$$

Then we take

$$q^* = \begin{cases} -\nabla u_0 & \text{in } \Omega_0, \\ -\left(\dfrac{\partial V}{\partial y}, -\dfrac{\partial V}{\partial x}\right)^{\mathrm{T}} & \text{in } \Omega \backslash \Omega_0. \end{cases} \tag{4.6}$$

This function is admissible, since $\forall v \in H_0^1(\Omega)$, we have

$$\begin{aligned} \langle \operatorname{div} q^*, v \rangle &= -\int_{\Omega} q^* \nabla v \, dx \\ &= \int_{\Omega_0} \nabla u_0 \nabla v \, dx + \int_{\Omega \backslash \Omega_0} \left(\frac{\partial V}{\partial y}, -\frac{\partial V}{\partial x}\right) \nabla v \, dx \\ &= \int_{\Gamma_0} -\frac{\partial u_0}{\partial n_0} v \, ds + \int_{\Omega_0} f v \, dx + \int_{\Gamma_0} \frac{\partial V}{\partial s} v \, ds \\ &= \int_{\Omega_0} f v \, dx \\ &= \int_{\Omega} f v \, dx. \end{aligned}$$

The chosen q^* has the property that $|q^*(x)| \leqslant \sqrt{\xi_0}$. The inequality is obvious for

$x \in \Omega_0$. For $x \in \Omega \backslash \Omega_0$, we have

$$|q^*|^2 = |\nabla V|^2 = \left|\frac{\partial V}{\partial r}\right|^2 + \frac{1}{r^2}\left|\frac{\partial V}{\partial \theta}\right|^2 = \left(\int_{\omega/2}^{\theta} \frac{\partial u_0}{\partial n_0}\bigg|_{\Gamma_0} \mathrm{d}\theta\right)^2 + \left(\frac{\partial u_0}{\partial n_0}\bigg|_{\Gamma_0}\right)^2$$
$$\leqslant \frac{\xi_0}{(\omega/2)^2 + 1}\left[\left(\theta - \frac{\omega}{2}\right)^2 + 1\right] \leqslant \xi_0.$$

Now, the estimates (3.9) and (3.11) give

$$J(u_0, \Lambda u_0) - J(u, \Lambda u) \leqslant \int_{\Omega\backslash\Omega_0} \left\{ \int_{|q^*|^2}^{|\nabla u_0|^2} \tfrac{1}{2} a(\xi)\,\mathrm{d}\xi + (q^* + \nabla u_0)\nabla g_1 + |q^*|^2 - |\nabla u_0|^2 \right\} \mathrm{d}x, \tag{4.7}$$

and,

$$\|\nabla(u - u_0)\|^2_{L^2(\Omega_1(u,u_0))} + \beta \|\nabla(u - u_0)\|^2_{L^2(\Omega\backslash\Omega_1(u,u_0))}$$
$$\leqslant \int_{\Omega\backslash\Omega_0} \left\{ \int_{|q^*|^2}^{|\nabla u_0|^2} a(\xi)\,\mathrm{d}\xi + 2(q^* + \nabla u_0)\nabla g_1 + 2(|q^*|^2 - |\nabla u_0|^2) \right\} \mathrm{d}x. \tag{4.8}$$

When the boundary condition around the corner is of Neumann type, we assume there exists an $r_0 > 0$ such that in $\Omega_0 = \Omega \cap \{r > r_0\}$, $|\nabla u_0| \leqslant \sqrt{\xi_0}/\sqrt{(\omega^2 + 1)}$. Again, denote $\Gamma_0 = \Omega \cap \{r = r_0\}$. Besides the assumption (4.4), we also assume that g_2 vanishes around O, i.e.

$$g_2 = 0 \quad \text{on } \Gamma_2 \cap \partial(\Omega\backslash\Omega_0).$$

Then we take

$$q^* = \begin{cases} -\nabla u_0 & \text{in } \Omega_0, \\ -\left(\dfrac{\partial V}{\partial y}, -\dfrac{\partial V}{\partial x}\right)^T & \text{in } \Omega\backslash\Omega_0, \end{cases} \tag{4.9}$$

where, instead of (4.5),

$$V(r, \theta) = r \int_0^{\theta} \frac{\partial u_0}{\partial n_0}\,\mathrm{d}\theta, \quad 0 \leqslant \theta \leqslant \omega,\ 0 < r \leqslant r_0.$$

From the linear problem for u_0 and the assumptions made on f and g_2, we have the equality

$$\int_0^{\omega} \frac{\partial u_0}{\partial n_0}\bigg|_{\Gamma_0} \mathrm{d}\theta = 0,$$

hence, q^* is an admissible function. And once more, $|q^*(x)| \leqslant \sqrt{\xi_0}$. Therefore, the estimates (4.7) and (4.8) hold.

The situation where the boundary condition around the corner is of mixed type can be treated similarly.

5. *A posteriori* error estimates for linearization (discrete problems)

Let $S \subset V$ be a finite element subspace based on some triangulation of the domain Ω and some selection of finite element degrees. For simplicity, we assume the Dirichlet boundary condition on Γ_1 can be represented exactly as the trace on Γ_1 of some function in S, i.e. we assume $g_1 \in S$. We denote the finite element subset

$$S_{\Gamma_1, g_1} = \{v \in S \mid v = g_1 \text{ on } \Gamma_1\}$$

and the finite element subspace

$$S_{\Gamma_1} = \{v \in S \mid v = 0 \text{ on } \Gamma_1\}.$$

The corresponding finite element solution of the non-linear problem (3.6) solves:

$$u^{\mathrm{FE}} \in S_{\Gamma_1, g_1}, \quad \int_\Omega a(|\nabla u^{\mathrm{FE}}|^2)\nabla u^{\mathrm{FE}} \nabla v \, \mathrm{d}x = \int_\Omega fv \, \mathrm{d}x + \int_{\Gamma_2} g_2 v \, \mathrm{d}s, \ \forall v \in S_{\Gamma_1}, \quad (5.1)$$

and the corresponding finite element solution of the linearized problem (3.7) solves:

$$u_0^{\mathrm{FE}} \in S_{\Gamma_1, g_1}, \quad \int_\Omega \nabla u_0^{\mathrm{FE}} \nabla v \, \mathrm{d}x = \int_\Omega fv \, \mathrm{d}x + \int_{\Gamma_2} g_2 v \, \mathrm{d}s, \quad \forall v \in S_{\Gamma_1}. \quad (5.2)$$

Our task of the section is to give computable bounds on the error $u^{\mathrm{FE}} - u_0^{\mathrm{FE}}$ by using u_0^{FE} only. To do this, we apply the duality theory presented in section 2. We take

$V = S$, with $H^1(\Omega)$ – norm and $V^* = S$ algebraically;

$Q = Q^* = (L^2(\Omega))^2 \times L^2(\Gamma_2)$;

$\Lambda v = (\nabla v, v|_{\Gamma_2})$.

We easily see that the operator Λ^* from Q^* to V^* is uniquely determined by the relation:

$$(\Lambda^* q^*, v) = \int_\Omega \nabla v q_1^* \, \mathrm{d}x + \int_{\Gamma_2} v q_2^* \, \mathrm{d}s, \quad \forall v \in S, \ q^* = (q_1^*, q_2^*) \in Q^*,$$

and that $\Lambda^* \in \mathscr{L}(Q^*, V^*)$. As in section 3, we define the energy function on $V = S$:

$$J(v, \Lambda v) = F(v) + G(\Lambda v),$$

where

$$F(v) = \begin{cases} -\int_\Omega fv \, \mathrm{d}x & \text{if } v \in S_{\Gamma_1, g_1}, \\ +\infty & \text{otherwise;} \end{cases}$$

$$G(\Lambda v) = \int_\Omega \int_0^{|\nabla v|^2} \frac{1}{2} a(\xi) \, \mathrm{d}\xi \, \mathrm{d}x - \int_{\Gamma_2} g_2 v \, \mathrm{d}s.$$

The finite element solution u^{FE} is also the unique solution of the minimization problem

$$\inf_{v \in S} J(v, \Lambda v).$$

Now, exactly as in section 3.2, we can prove the following error estimate:

$$\|\nabla(u^{\mathrm{FE}} - u_0^{\mathrm{FE}})\|^2_{L^2(\Omega_1(u^{\mathrm{FE}},u_0^{\mathrm{FE}}))} + \beta\|\nabla(u^{\mathrm{FE}} - u_0^{\mathrm{FE}})\|^2_{L^2(\Omega\backslash\Omega_1(u^{\mathrm{FE}},u_0^{\mathrm{FE}}))}$$

$$\leqslant \int_\Omega \int_{b(|q^*|)^2}^{|\nabla u_0^{\mathrm{FE}}|^2} a(\xi)\,\mathrm{d}\xi\,\mathrm{d}x + 2\int_\Omega ((q^* + \nabla u_0^{\mathrm{FE}})\nabla g_1 + |q^*|\,b(|q^*|) - |\nabla u_0^{\mathrm{FE}}|^2)\,\mathrm{d}x$$

$$\forall q^* \in (L^2(\Omega))^2,\ -\int_\Omega q^*\nabla v\,\mathrm{d}x = \int_\Omega fv\,\mathrm{d}x + \int_{\Gamma_2} g_2 v\,\mathrm{d}s,\quad \forall v \in S_{\Gamma_1}, \tag{5.3}$$

where

$$\Omega_1(u^{\mathrm{FE}}, u_0^{\mathrm{FE}}) = \{x \in \Omega \mid |\nabla u^{\mathrm{FE}}(x)|,\ |\nabla u_0^{\mathrm{FE}}(x)| \leqslant \sqrt{\xi_0}\}.$$

To use (5.3) to get a good error bound on $u^{\mathrm{FE}} - u_0^{\mathrm{FE}}$, we need to make some suitable selection of the auxiliary function q^*. Comparing the constraint on the auxiliary function q^* with the finite element system (5.2), we see that

$$q^* = -\nabla u_0^{\mathrm{FE}} \tag{5.4}$$

is an admissible function. With the simple selection (5.4), we get from (5.3) that

$$\|\nabla(u^{\mathrm{FE}} - u_0^{\mathrm{FE}})\|^2_{L^2(\Omega_1(u^{\mathrm{FE}},u_0^{\mathrm{FE}}))} + \beta\|\nabla(u^{\mathrm{FE}} - u_0^{\mathrm{FE}})\|^2_{L^2(\Omega\backslash\Omega_1(u^{\mathrm{FE}},u_0^{\mathrm{FE}}))}$$

$$\leqslant \frac{(1-\beta)^2}{\beta}\int_{\Omega_n(u_0^{\mathrm{FE}})} (|\nabla u_0^{\mathrm{FE}}| - \sqrt{\xi_0})^2\,\mathrm{d}x, \tag{5.5}$$

where

$$\Omega_n(u_0^{\mathrm{FE}}) = \{x \in \Omega \mid \nabla u_0^{\mathrm{FE}}(x)| > \sqrt{\xi_0}\}.$$

If $\beta > 0$ is not close to 0, (5.5) will produce a good error bound (at least when finite element solutions are sufficiently close to the solutions of the continuous problems, cf. Remark 6.3). If β is close to 0, however, the estimate (5.5) will give useless error bound, because of the factor $1/\beta$. Thus, an alternative selection of the auxiliary function q^* is needed. Once more, let us have the situation as described in Selection 2 of section 4.

Take an $r_0^{\mathrm{FE}} > 0$ such that in $\Omega_0^{\mathrm{FE}} = \Omega \cap \{r > r_0^{\mathrm{FE}}\}$, $|\nabla u_0^{\mathrm{FE}}| \leqslant \sqrt{\xi_0}/\sqrt{[(\omega/2)^2 + 1]}$. Denote $\Gamma_0^{\mathrm{FE}} = \Omega \cap \{r = r_0^{\mathrm{FE}}\}$, n_0 the unit normal vector on Γ_0^{FE} in the direction of being away from O. We assume

$$f(x) = 0. \quad x \in \Omega\backslash\Omega_0^{\mathrm{FE}},$$

and define, in $\Omega\backslash\Omega_0^{\mathrm{FE}}$, a function

$$V(r, \theta) = r\int_{\omega/2}^{\theta} \left.\frac{\partial u_0^{\mathrm{FE}}}{\partial n_0}\right|_{\Gamma_0} \mathrm{d}\theta,\quad 0 \leqslant \theta \leqslant \omega,\ 0 < r \leqslant r_0.$$

Then it can be proved that

$$q^* = \begin{cases} -\nabla u_0^{\mathrm{FE}} & \text{in } \Omega_0^{\mathrm{FE}} \\ -\left(\dfrac{\partial V}{\partial y}, -\dfrac{\partial V}{\partial x}\right)^T & \text{in } \Omega\backslash\Omega_0^{\mathrm{FE}}, \end{cases} \tag{5.6}$$

is an admissible function satisfying $|q^*(x)| \leqslant \sqrt{\xi_0}$. With the selection (5.6), we obtain the following computable error estimate which is insensitive to β as β approaches 0:

$$\|\nabla(u^{FE} - u_0^{FE})\|^2_{L^2(\Omega_1(u^{FE},u_0^{FE}))} + \beta\|\nabla(u^{FE} - u_0^{FE})\|^2_{L^2(\Omega\setminus\Omega_1(u^{FE},u_0^{FE}))}$$

$$\leqslant \int_{\Omega\setminus\Omega_0^{FE}} \left\{ \int_{|q^*|^2}^{|\nabla u_0^{FE}|^2} a(\xi)\,d\xi + 2(q^* + \nabla u_0^{FE})\nabla g_1 + 2(|q^*|^2 - |\nabla u_0^{FE}|^2) \right\} dx. \tag{5.7}$$

6. Numerical experiments

We present two numerical examples to show the use of the estimates (4.2), (4.3), (4.7) and (4.8). We will compare the estimates resulted from the two different selections of the auxiliary function q^*. For the first example, we know the exact solutions for both the non-linear problem and the linear problem. Hence, we will see how effective our estimates are by comparing our error bounds with exact errors.

Experiment 6.1. We take Ω a unit disk excluding a small hole:

$$\Omega = \{(r, \theta) \mid r_* < r < 1\}, \tag{6.1}$$

where $r_* > 0$ is a small number. Let us consider a family of non-linear problems with a small parameter $\lambda > 0$:

$$\begin{aligned} -\operatorname{div}(a(|\nabla u|^2)\nabla u &= 0 && \text{in } \Omega, \\ u &= \lambda \log(1/r_*) && \text{for } r = 1, \\ u &= 0 && \text{for } r = r_*, \end{aligned} \tag{6.2}$$

where a is the function defined in (3.1). The exact solution is

$$u = \begin{cases} \sqrt{\xi_0} r_1 \log r + \lambda \log \dfrac{1}{r_*} & \text{for } r_1 < r < 1, \\ \dfrac{\sqrt{\xi_0}}{\beta}\left(r_1 \log \dfrac{r}{r_*} - (1-\beta)(r - r_*) \right) & \text{for } r_* < r < r_1, \end{cases} \tag{6.3}$$

where r_1 is the unique positive solution of the equation:

$$\log r_* \cdot r_1 - (1-\beta)(\log r_1 - 1) r_1 = (1-\beta) r_* + \beta \log r_* \cdot r_0 \tag{6.4}$$

with

$$r_0 = \lambda / \sqrt{\xi_0}. \tag{6.5}$$

The linearized problem:

$$\begin{aligned} -\Delta u_0 &= 0 && \text{in } \Omega, \\ u &= \lambda \log(1/r_*) && \text{for } r = 1, \\ u_0 &= 0 && \text{for } r = r_*, \end{aligned} \tag{6.6}$$

has the solution:

$$u_0 = \lambda \log \frac{r}{r_*}. \tag{6.7}$$

It is easily verified that $r_1 < r_0$. Therefore,

$$\Omega_1(u_0) = \{(r,\theta)\,|\,r_0 < r < 1\} \subset \Omega_1(u) = \{(r,\theta)\,|\,r_1 < r < 1\}.$$

Thus, in some sense, the linearized problem is more rigid than the non-linear problem.

We take $\xi_0 = 1$, $r_* = 0.01$. For various values of β, λ, we compute the following quantities:

(a) $r_0 = \lambda/\sqrt{\xi_0} = \lambda$.

(b) r_1, determined from equation (6.4).

(c) The square root of the (non-linear) energy of the linear solution:

$$\sqrt{[J(u_0,\Lambda u_0)]} = \sqrt{\pi}\lambda\left\{(1-\beta)\left[\frac{3}{2} - 2\frac{r_*}{r_0} + \frac{1}{2}\left(\frac{r_*}{r_0}\right)^2\right] + \log\frac{1}{r_0} + \beta\log\frac{r_0}{r_*}\right\}^{1/2}.$$

(d) The square root of the (non-linear) energy of the non-linear solution:

$$\sqrt{[J(u,\Lambda u)]} = \sqrt{(\pi\xi_0)}r_1\left\{\log\frac{1}{r_1} + \frac{1}{\beta}\log\frac{r_1}{r_*} - \frac{1-\beta}{2\beta}\left(1 - \left(\frac{r_*}{r_1}\right)\right)\right\}^{1/2}.$$

(e) The square root of the difference in energy:

$$E_J(u,u_0) = \sqrt{[J(u_0,\Lambda u_0) - J(u,\Lambda u)]}.$$

(f) The error in 'energy norm':

$$E(u,u_0) \equiv \sqrt{[\tfrac{1}{2}\|\nabla(u-u_0)\|^2_{L^2(\Omega_1(u,u_0))} + \frac{\beta}{2}\|\nabla(u-u_0)\|^2_{L^2(\Omega\setminus\Omega_1(u,u_0))}]}$$

$$= \sqrt{(\pi\xi_0)}\left\{(r_0-r_1)^2\left(\log\frac{1}{r_0} + \beta\log\frac{r_0}{r_1}\right)\right.$$

$$+\frac{1}{\beta}\left[(r_1-\beta r_0)^2\log\frac{r_1}{r_*} - 2(1-\beta)(r_1-\beta r_0)(r_1-r_*)\right.$$

$$\left.\left.+\frac{1}{2}(1-\beta)^2(r_1^2-r_*^2)\right]\right\}^{1/2}.$$

(g) An error estimate resulted from Selection 1. We take

$$q^* = -\nabla u_0,$$

then an upper bound for $E_J(u,u_0)$ and $E(u,u_0)$ is (cf. (4.2) and (4.3)):

$$\mathbf{E}sted_1 = (1-\beta)\lambda\sqrt{\frac{\pi}{\beta}}\left\{\log\frac{r_0}{r_*} + 2\frac{r_*}{r_0} - \frac{1}{2}\left(\frac{r_*}{r_0}\right)^2 - \frac{3}{2}\right\}^{1/2}.$$

(h) An error estimate resulted from Selection 2. We take

$$q^* = \begin{cases} -\dfrac{\lambda}{r}\begin{pmatrix}\cos\theta\\ \sin\theta\end{pmatrix}, & r_2 < r < 1\\[2ex] -\dfrac{\lambda}{r_2}\begin{pmatrix}(\theta-\pi)\sin\theta + \cos\theta\\ -(\theta-\pi)\cos\theta + \sin\theta\end{pmatrix}, & r_* < r < r_2,\end{cases} \tag{6.8}$$

where

$$r_2 = r_0\sqrt{\pi^2 + 1}.$$

Then an upper bound for $E_J(u, u_0)$ and $E(u, u_0)$ is (cf. (4.7), (4.8)):

$$\text{E}sted_2 = \lambda\sqrt{\pi}\left\{\frac{\pi^2}{6} - \frac{3}{2}\beta + \log\frac{r_2}{r_0} + 2\frac{r_*}{r_2} - 2(1-\beta)\frac{r_*}{r_0} - \frac{1}{2}\left(\frac{\pi^2}{3} + 1\right)\left(\frac{r_*}{r_2}\right)^2 + \frac{1}{2}(1-\beta)\left(\frac{r_*}{r_0}\right)^2\right\}^{1/2}.$$

Tables 1–3 show the numerical results. We denote $J_0 = J(u_0, \Lambda u_0)$, $J = J(u, \Lambda u)$.

We see that min$\{\text{E}sted_1, \text{E}sted_2\}$ is a very good error bound for any values of β and λ. For β away from 0, E$sted_1$ provides a better bound, while when β is close to 0, we should use E$sted_2$ as our error bound.

Experiment 6.2. Let Ω be a circular corner domain:

$$\Omega = \{(r, \theta) \mid 0 < r < 1, 0 < \theta < \omega\}, \quad \omega > \pi. \tag{6.9}$$

We use again the notation $\alpha = \pi/\omega$. Consider a particular case when the linear solution is

$$u_0 = \lambda r^\alpha \sin\alpha\theta, \tag{6.10}$$

Table 1. Exact errors and estimated error bounds ($\lambda = 0.1$)

β	r_1	$\sqrt{J_0}$	$\sqrt{J}$	$E_J(u, u_0)$	$E(u, u_0)$	E$sted_1$	E$sted_2$
0.9	0.97^{-1}	0.3762^{+0}	0.3760^{+0}	0.1383^{-1}	0.1383^{-1}	0.1866^{-1}	0.3367^{+0}
0.5	0.79^{-1}	0.3592^{+0}	0.3498^{+0}	0.8141^{-1}	0.8080^{-1}	0.1252^{+0}	0.3171^{+0}
0.1	0.43^{-1}	0.3413^{+0}	0.2774^{+0}	0.1988^{+0}	0.1867^{+0}	0.5038^{+0}	0.2967^{+0}
0.01	0.20^{-1}	0.3371^{+0}	0.2088^{+0}	0.2646^{+0}	0.2304^{+0}	0.1753^{+1}	0.2919^{+0}
0.001	0.13^{-1}	0.3367^{+0}	0.1804^{+0}	0.2843^{+0}	0.2391^{+0}	0.5593^{+1}	0.2914^{+0}

Table 2. Exact errors and estimated error bounds ($\lambda = 0.015$)

β	r_1	$\sqrt{J_0}$	$\sqrt{J}$	$E_J(u, u_0)$	$E(u, u_0)$	E$sted_1$	E$sted_2$
0.9	0.15^{-1}	0.5704^{-1}	0.5704^{-1}	0.3470^{-3}	0.3470^{-3}	0.3608^{-3}	0.3814^{-1}
0.5	0.15^{-1}	0.5700^{-1}	0.5696^{-1}	0.2264^{-2}	0.2263^{-2}	0.2420^{-2}	0.3808^{-1}
0.1	0.14^{-1}	0.5696^{-1}	0.5645^{-1}	0.7595^{-2}	0.7579^{-2}	0.9742^{-2}	0.3802^{-1}
0.01	0.12^{-1}	0.5695^{-1}	0.5522^{-1}	0.1394^{-1}	0.1376^{-1}	0.3389^{-1}	0.3801^{-1}
0.001	0.11^{-1}	0.5695^{-1}	0.5434^{-1}	0.1704^{-1}	0.1665^{-1}	0.1081^{+0}	0.3801^{-1}

Table 3. Exact errors and estimated error bounds ($\lambda = 0.0101$)

β	r_1	$\sqrt{J_0}$	$\sqrt{J}$	$E_J(u, u_0)$	$E(u, u_0)$	E$sted_1$	E$sted_2$
0.9	0.10^{-1}	0.3842^{-1}	0.3842^{-1}	0.1076^{-5}	0.1076^{-5}	0.1077^{-5}	0.2372^{-1}
0.5	0.10^{-1}	0.3842^{-1}	0.3842^{-1}	0.7215^{-5}	0.7215^{-5}	0.7227^{-5}	0.2372^{-1}
0.1	0.10^{-1}	0.3842^{-1}	0.3842^{-1}	0.2886^{-4}	0.2886^{-4}	0.2909^{-4}	0.2372^{-1}
0.01	0.10^{-1}	0.3842^{-1}	0.3842^{-1}	0.9414^{-4}	0.9414^{-4}	0.1012^{-3}	0.2372^{-1}
0.001	0.10^{-1}	0.3842^{-1}	0.3842^{-1}	0.2135^{-3}	0.2135^{-3}	0.3229^{-3}	0.2372^{-1}

where $\lambda > 0$ is a parameter. Since $|\nabla u_0| = \lambda \alpha r^{\alpha-1}$, the linear region of u_0 is

$$\Omega_l(u_0) = \{x \in \Omega \mid r \geqslant r_0\},$$

where the radius

$$r_0 = \left(\frac{\lambda\alpha}{\sqrt{\xi_0}}\right)^{1/(1-\alpha)}.$$

Let us estimate the difference between u_0 and u, the solution of the non-linear problem:

$$\begin{aligned} -\operatorname{div}(a(|\nabla u|^2)\nabla u) &= 0 \quad \text{in } \Omega, \\ u &= \lambda r^\alpha \sin\alpha\theta \quad \text{on } \partial\Omega. \end{aligned} \tag{6.11}$$

We try to give an upper bound for both $E_J(u, u_0)$ and $E(u, u_0)$, which were defined in Experiment 6.1. If we use Selection 1

$$q^* = -\nabla u_0,$$

then an upper bound for the error $u - u_0$ is

$$\mathrm{E}sted_1 = \lambda(1-\beta)\sqrt{\frac{\pi}{\beta}}\,\xi_0^{-\alpha/2(1-\alpha)}\left\{\frac{1}{4}\alpha^{2\alpha/(1-\alpha)} - \left(\frac{1}{1+\alpha} - \frac{1}{4}\right)\alpha^{(1+\alpha)/(1-\alpha)}\right\}^{1/2}.$$

If we use Selection 2

$$q^* = \begin{cases} -\nabla u_0 & \text{in } \Omega_1, \\ -\left(\dfrac{\partial V}{\partial y}, -\dfrac{\partial V}{\partial x}\right)^{\mathrm{T}} & \text{in } \Omega\backslash\Omega_1, \end{cases} \tag{6.12}$$

where

$$\Omega_1 = \{x \in \Omega \mid r > r_1\},$$

$$r_1 = \left(\frac{\lambda}{\sqrt{\xi_0}}\right)^{1/(1-\alpha)},$$

and

$$V(r,\theta) = -\alpha r_1^{\alpha-1} r \cos\alpha\theta \quad \text{in } \Omega_1,$$

then an upper bound for the error is

$$\begin{aligned} \mathrm{E}sted_2 = {} & \lambda^{1/(1-\alpha)}\sqrt{\pi}\,\xi_0^{-\alpha/2(1-\alpha)} \\ & \times\left\{(1-\beta)\left(-\frac{1}{4}\alpha^{2\alpha/(1-\alpha)} + \left(\frac{1}{1+\alpha} - \frac{1}{4}\right)\alpha^{(1+\alpha)/(1-\alpha)}\right) + \frac{(1-\alpha)^2}{8\alpha}\right\}^{1/2}. \end{aligned}$$

We take $\xi_0 = 1$, and $\omega = 3\pi/2$ (i.e. Ω is an L-shape domain), the numerical results are shown in Table 4.

We see, as in the last example, $\mathrm{E}sted_1$ provides a better bound for β away from 0, while $\mathrm{E}sted_2$ is better than $\mathrm{E}sted_1$ when β is close to 0.

Remark 6.3. We have seen in the numerical examples that our *a posteriori* error bounds for the continuous problems are effective. When β is not close to 0, the error

Table 4. Estimated error bounds for problems on an L-shape domain

λ	β	$\sqrt{[J(u_0, \nabla u_0)]}$	$Ested_1$	$Ested_2$
0.4	0.9	0.3549^{+0}	0.6861^{-3}	0.1624^{-1}
	0.5	0.3545^{+0}	0.4602^{-2}	0.1571^{-1}
	0.1	0.3544^{+0}	0.1852^{-1}	0.1516^{-1}
	0.01	0.3544^{+0}	0.6444^{-1}	0.1504^{-1}
	0.001	0.3544^{+0}	0.2056^{+0}	0.1503^{-1}
0.01	0.9	0.8862^{-2}	0.1072^{-7}	0.2538^{-6}
	0.5	0.8862^{-2}	0.7191^{-7}	0.2455^{-6}
	0.1	0.8862^{-2}	0.2894^{-6}	0.2369^{-6}
	0.01	0.8862^{-2}	0.1007^{-5}	0.2350^{-6}
	0.001	0.8862^{-2}	0.3213^{-5}	0.2348^{-6}

bound from the simple Selection 1 is better than that from the Selection 2. On the other hand, when β is close to 0, the error bound from the Selection 2 is better than that from the Selection 1. By continuity, at least when finite element solutions are sufficient close to the exact solutions, we can claim that our *a posteriori* error bounds for the discrete problems are effective, too. When β is not close to 0, the error bound (5.5) is better than (5.7), while when β is close to 0, the error bound (5.7) is better than (5.5).

7. Some Remarks

Remarks 7.1. In the above discussions, we assumed a concrete form for the non-linear problem. This is only for the purpose of making specific computations and of showing the effectiveness of the error estimates. In fact, we do not depend on a concrete form for the non-linear problem. To be able to provide error estimates for the effects of mathematical idealizations on solutions, we only need a description on the range of the coefficient function. This is a crucial point which allows the useful applications of the suggested *a posteriori* error estimates on practical problems, where all the data contain certain degree of uncertainty. Of course, the more accurate the description of the non-linear problem, the more accurate the error estimates.

Remark 7.2. In the above discussions, we considered the effect of the linearization. The mathematical idealizations of a practical problem can be more complicated than the linearization, and the general framework for deriving *a posteriori* error estimates discussed earlier allows us to get similar estimates.

Remark 7.3. The sample problem discussed above is an elliptic boundary value problem for a Laplacian-like differential operator. The idea for deriving *a posteriori* error estimates works for more complicated problems commonly seen in mechanics.

Remark 7.4. In fluid mechanics, most problems are time-dependent. Derivation of *a posteriori* error estimates for the effects of mathematical idealizations on solutions is more involved. We will present related results in a future paper.

Remark 7.5. *A posteriori* error estimates can be used to devise some adaptive numerical procedures. In [9], such estimates are given for the regularization of non-differentiable minimization problems. As another application in numerical procedures, we will study iterative procedures for solving non-linear problems. *A posteriori* error estimates will provide us a convenient and efficient stopping criterion.

Acknowledgement

The author would like to thank Professor Ivo Babuška for his advice, valuable discussions, encouragement and support. He also thanks Professor Gilbert and two referees for very helpful comments.

References

1. Arthurs A. M., *Complementary Variational Principles*, 2nd edn., Oxford Math. Monographs, Oxford, 1980.
2. Babuška, I., 'The problem of modelling the elastomechanics in engineering', *Comput. methods appl. mech. eng.*, **82**, 155–182 (1990).
3. Babuška, I. and Pitkäranta, J., 'The plate paradox for hard and soft simple support', *SIAM J. Math. Anal.*, **21**, 551–576 (1990).
4. Ekeland, I. and Temam, R., *Convex Analysis and Variational Problems*, North-Holland, Amsterdam, 1976.
5. Grisvard, P., *Elliptic Problems in Nonsmooth Domains*, Pitman, Boston, 1985.
6. Han, W., 'Quantitative error estimates for material idealization of torsion problems', *Math. Comput. Modelling Int. J.* **15**, No. 9, 47–54 (1991).
7. Han, W., 'A regularization procedure for a simplified friction problem', *Math. Comput. Modelling Int. J.* **15**, No. 8, 65–70 (1991).
8. Han, W., 'Quantitative error estimate in modelling the laminar stationary flow of a Bingham fluid', *Appl. Math. Comput.*, **47**, 15–24 (1992).
9. Han, W., 'Finite element analysis of a holonomic elastic–plastic problem', *Numer. Math.* **60**, 493–508 (1992).
10. Hlaváček, I., Haslinger, J., Nečas, J. and Lovíšek J., *Solution of Variational Inequalities in Mechanics*, Springer, New York, 1988.
11. Mikhlin, S. G., *Variatonal Methods in Mathematical Physics*, Pergamon Press, Oxford, 1964.
12. Noble, B. and Sewell, M. J., 'On dual extremum principles in applied mathematics', *J. Inst. Math. Appl.*, **9**, 123–193 (1972).
13. Sewell, M. J., 'Dual approximation principles and optimization in continuum mechanics', *Phil. Trans. Roy. Soc.* (London) **A265**, 319–351 (1969).
14. Velte, W., 'Complementary variational principles', in E. Stein and W. Wendland (eds), *Finite Element and Boundary Element Techniques from Mathematical and Engineering Point of View*, Springer, Wien-New York 1988.