UNIVERSITAS COMENIANA
ACTA MATHEMATICA UNIVERSITATIS COMENIANAE
LVIII—LIX

A POSTERIORI ERROR ESTIMATE OF APPROXIMATE SOLUTIONS TO A NONLINEAR ELIPTIC BOUNDARY VALUE PROBLEM

JURAJ WEISZ, Bratislava

Abstract. The paper deals with a computable a posteriori error estimate of approximate solutions to the nonlinear elliptic boundary value problem with the Dirichlet boundary condition

$$-\sum_{i=1}^{2} \frac{\partial}{\partial x_i}(a_i(x, \nabla u)) + g(x, u) = f(x) \text{ in } \Omega,$$

$$u|_{\partial\Omega} = 0.$$

The convergence of the presented error estimate to the true error is proved. The computation of the error estimate requires the computation of a finite system of linear algebraic equations in each step.

0. Introduction

This paper deals with an a posteriori error estimate of the error of the approximate solution to the nonlinear elliptic boundary value problem with the homogeneous Dirichlet boundary condition

$$-\sum_{i=1}^{2} \frac{\partial}{\partial x_i} a_i(x, \nabla u) + g(x, u) = f(x) \text{ in } \Omega, \qquad u|_{\partial\Omega} = 0 \tag{1}$$

The linear cases $a_i(x, t_1, t_2) = \sum_{j=1}^{2} a_{ij}(x)\, t_j$, $i = 1, 2$, $g(x, t) = 0$ and $g(x, t) = a_0(x)\, t$, $a_0 > 0$ for two-dimensional problems have been studied in [4], [5], [8]. A special nonlinear case (with linear main part) for one-dimensional problems has been studied in [7]. However the method of [7] cannot be straight generalized for two-dimensional problems. The mildly nonlinear case $a_i(x, t_1,$

$t_2) = t_i$, $i = 1, 2$, $g(x, t) = g(t)$ for two-dimensional problems has been studied in [9]. This paper deals with the general case. We shall show that the presented error bound can be made arbitrarily small for a sufficiently good approximation of the exact solution. Our error bound can be computed by solving a finite system of nonlinear (Theorem 2) or linear (Theorem 3) equations.

The paper consists of four sections. The first section contains basic notation. The aim of the second section is to prove Proposition 1, which is the basis of our error estimates. The third section deals with the construction of the dual problem. Our a posteriori error estimates (based on methods for approximate solution of the the dual problem) are derived in the last section of the paper.

1. Notation

In the sequel we shall adopt the following notation: If B is a Banach space, B' denotes its dual and $\langle \cdot, \cdot \rangle_B$ denotes the duality pairing between B' and B. If $\mathscr{B}$: $B \to \bar{R}$ is a functional then $\mathscr{B}^*$: $B' \to \bar{R}$ denotes its conjugate functional

$$\mathscr{B}^*(b') = \sup_{b \in B} \{\langle b', b \rangle_B - \mathscr{B}(b)\}.$$

If B, C are Banach spaces then $L(B, C)$ denotes the space of all linear bounded operators from B to C with the usual topology and $A' \in L(C', B')$ denotes the transpose of $A \in L(B, C)$ defined by $\langle A'c', b \rangle_B = \langle c', Ab \rangle_C$ for $b \in B$, $c' \in C'$. The norm in $L(B, C)$ will be denoted as usual $\| \cdot \|_{L(B, C)}$.

$\Omega \subset R^2$ denotes a simply connected bounded domain with polygonal boundary $\partial\Omega$, H denotes the Lebesgue space $L_2(\Omega)$, endowed with the scalar product $(u, v) = \int_\Omega uv \, dx$ and with the norm $\| u \|_{L_2} = (u, u)^{1/2}$.$\mathbf{H}$ denotes the space $L_2(\Omega) \times L_2(\Omega)$ endowed with the scalar product $[\mathbf{u}, \mathbf{v}] = [(u_1, u_2), (v_1, v_2)] = \int_\Omega u_1 v_1 + u_2 v_2 \, dx$ and with the norm $[\mathbf{u}] = [\mathbf{u}, \mathbf{u}]^{1/2}$. H and $\mathbf{H}$ are supposed to be identified with their duals (using the Riesz's representation).

V and U are subspaces of the Sobolev space $W^{1,2}(\Omega)$:

$V = W_0^{1,2}(\Omega) = \{v \in W^{1,2} = (\Omega) \mid v|_{\partial\Omega} = 0$ in the sense of traces$\}$,

$U = \{v \in W^{1,2}(\Omega) \mid \int_\Omega v \, dx = 0\}$.

Both U and V are endowed with the inner product

$$((u, v)) = \sum_{i=1}^{2} \int_\Omega \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_i} \, dx$$

and with the norm $\| u \| = ((u, u))^{1/2}$. U_n, $n = 1, 2, \ldots$ denotes a sequence of finitedimensional subspaces of U and $P_n \in L(U, U_n)$, $n = 1, 2, \ldots$ the sequence of corresponding orthogonal projectors. We suppose that the sequence U_n, $n = 1, 2, \ldots$ satisfies the properties

$$U_n \subset U_{n+1}, \; n = 1, 2, \ldots,$$

$$\forall v \in U: \lim_{n \to \infty} \| P_n v - v \| = 0.$$

Finally we give some definitions of operators: $\Lambda \in L(V, H)$, $\Lambda v = v$, $\nabla \in L(V, H)$, $\nabla v = \left(\frac{\partial v}{\partial x_1}, \frac{\partial v}{\partial x_2}\right)$, $L \in L(U, \mathbf{H})$, $Lv = \left(-\frac{\partial v}{\partial x_2}, \frac{\partial v}{\partial x_1}\right)$, $J \in L(U, U')$, $J = L'$, L, (J is the duality operator see e.g. [GGZ-Chapter III]), $I_n \in L(U_n, U)$, $I_n v = v$, $J_n \in L(U_n, U_n')$, $J_n = I_n' J I_n$, $L_n \in L(U_n, \mathbf{H})$, $L_n = L I_n$,

$R \in L(H, \mathbf{H})$, $Rl = -\left(\int_0^{x_1} \bar{l}(t, x_2)\, \mathrm{d}t, 0\right)$, where $\bar{l} = l$ in Ω, $\bar{l} = 0$ in $R^2 - \Omega$.

Clearly, $\langle \nabla' Rl, v \rangle_V = \langle \Lambda' l, v \rangle_V = \int_\Omega lv \, \mathrm{d}x$, $v \in V$.

2. Formulation of the problem

We suppose that $f \in L_2(\Omega)$ and that the functions $a_i: \Omega \times R^2 \to R$, $i = 1, 2$, $g: \Omega \times R \to R$ satisfy the Caratheodory conditions and the following growth conditions:

There exist $g_0 \in L_2(\Omega)$, β, $c > 0$ such that

$$|g(x, t)| \leq c(g_0(x) + |t|^\beta) \text{ for all } t \in R \text{ and for almost all } x \in \Omega,$$

there exist $g_i \in L_2(\Omega)$, $c_i > 0$ such that

$|a_i(x, t_1, t_2)| \leq c_i(g_i(x) + |t_1| + |t_2|)$ for all $(t_1, t_2) \in R^2$ and for almost all $x \in \Omega$, $i = 1, 2$.

We suppose that for almost all $x \in \Omega$ the function $g_x: R \to R$, $g_x(t) = g(x, t)$ is an increasing and surjective function of t, satisfying $g(x, 0) = 0$. The functions a_i, $i = 1, 2$ are supposed to be such that the operator $A_0: \mathbf{H} \to \mathbf{H}' = \mathbf{H}$

$$A_0(p_1, p_2) = (a_1(\cdot, p_1, p_2), a_2(\cdot, p_1, p_2))$$

is strongly monotone: there exists $m > 0$ such that

$$[A_0\mathbf{p} - A_0\mathbf{q}, \mathbf{p} - \mathbf{q}] \geq m[\mathbf{p} - \mathbf{q}]^2, \qquad \mathbf{p}, \mathbf{q} \in \mathbf{H},$$

lipschitzian: there exists $M > 0$ such that

$$[A_0\mathbf{p} - A_0\mathbf{q}]_* \leq M[\mathbf{p} - \mathbf{q}] \qquad \mathbf{p}, \mathbf{q} \in \mathbf{H},$$

and potential:

A_0 is a G-derivative of some functional $\mathscr{F}_0$.

Remark 1. Conditions which guarantee properties formulated above can be found e.g. in [3—Chapter III].

The operator T: $V \to V'$,

$$\langle Tw, v\rangle_V = \langle \nabla' A_0 \nabla w, v\rangle_V + \int_\Omega g(x, w(x))\, v(x)\, \mathrm{d}x$$

is strongly monotone (and thus coercitive) and potential and thus [3—Theorem III..4.1.] there exists solution to problem

$$u \in V,$$

$$\forall v \in V\colon \langle Tu, v\rangle_V = \sum_{i=1}^{2} \int_\Omega a_i(x, \nabla u) \frac{\partial v}{\partial x_i}\, \mathrm{d}x + \int_\Omega g(x, u)\, v\, \mathrm{d}x = \int_\Omega fv\, \mathrm{d}x. \tag{2}$$

(u is the weak solution to problem (1)). The solution is unique since T is strongly monotone. This solution is characterized by

$$u \in V, \qquad \mathscr{F}(u) \leq \mathscr{F}(v), \qquad \forall v \in V,$$

where $\mathscr{F}$: $V \to R$ is the potential of problem (1) defined by

$$\mathscr{F}(v) = \mathscr{F}_0(\nabla v) + j(v) - \int_\Omega fv\, \mathrm{d}x, \tag{3}$$

where

$$\mathscr{F}_0\colon \mathbf{H} \to R,\ \mathscr{F}_0(\mathbf{p}) = \int_0^1 [A_0 t\mathbf{p}, \mathbf{p}]\, \mathrm{d}t,$$

$$j\colon\ V \to R,\ j(v) = \int_\Omega \int_0^{v(x)} g(x, t)\, \mathrm{d}t\, \mathrm{d}x.$$

Both $\mathscr{F}_0$ and j are convex continuous functionals with G-derivatives $\mathscr{F}_0'$, j'

$$[\mathscr{F}_0', \mathbf{p}, \mathbf{q}] = [A_0\mathbf{p}, \mathbf{q}], \qquad \langle j'(v), w\rangle_v = \int_\Omega g(\cdot, v)\, w\, \mathrm{d}x.$$

Remark 2. A minimizing sequence of $\mathscr{F}$, ie. a sequence $u_n \in V$, $n = 1, 2, \ldots$ with the property $u_n \to u$ in V, can be constructed e.g. by the Ritz method (see [3 — Theorem III.4.3.]). Our aim is to construct a computable upper bound of

$\| u_n - u \|$. It is natural to require that the constructed error bound tends to zero for any minimizing sequence of $\mathscr{F}$. The following proposition is the basis for such estimates.

Proposition 1. Let u, $\mathscr{F}$ be defined by (2), (3). Then

$$\forall v \in V\colon \ \| u - v \|^2 \leq \frac{2}{m} (\mathscr{F}(v) - \mathscr{F}(u)).$$

Proof. From (2) and from [3-Lemma III.4.11] it follows that u is the unique minimizing point of the functional

$$\mathscr{F}_2(v) = \mathscr{F}_0(\nabla v) - \langle \Lambda'(f - g(\cdot, u)), v \rangle.$$

The convexity of j implies

$$\mathscr{F}(v) - \mathscr{F}(u) = \mathscr{F}_0(\nabla v) + j(v) - \langle \Lambda' f, v \rangle - \mathscr{F}_0(\nabla u) - j(u) + \langle \Lambda' f, u \rangle \geq$$

$$\geq \mathscr{F}_0(\nabla v) - \mathscr{F}_0(\nabla u) + \int_\Omega g(\cdot, u)(v - u)\,\mathrm{d}x + \langle \Lambda' f, u - v \rangle =$$

$$= \mathscr{F}_0(\nabla v) - \langle \Lambda' f - \Lambda' g(\cdot, u), v \rangle - \mathscr{F}_0(\nabla u) + \langle \Lambda' f - \Lambda' g(\cdot, u), u \rangle =$$

$$= \mathscr{F}_2(\nabla v) - \mathscr{F}_2(\nabla u)$$

and from [3-Proof of Theorem III.4.11] we have

$$\mathscr{F}_2(v) - \mathscr{F}_2(u) \geq \frac{m}{2} \| u - v \|^2. \ \blacksquare$$

Corollary 1. Let u, $\mathscr{F}$ be defined by (2), (3) and $u_n \in V$, $n = 1, 2, \ldots$ be a minimizing sequence of $\mathscr{F}$. Let $d_n \in R$, $n = 1, 2, \ldots$ be a sequence of real numbers satisfying the property

$$\begin{gathered} d_n \leq \mathscr{F}(u),\ n = 1, 2, \ldots, \\ d_n \to \mathscr{F}(u). \end{gathered} \tag{4}$$

Then

$$\| u_n - u \|^2 \leq \frac{2}{m} (\mathscr{F}(u_n) - d_n) \to 0.$$

Proof. It suffices to use Proposition 1, the continuity of $\mathscr{F}$ and (4). ■

As mentioned in Remark 2, a minimizing sequence of $\mathscr{F}$ can be constructed. Our problem is to construct a sequence $d_n \in R$, $n = 1, 2, \ldots$ satisfying (4).

3. Dual problem

In this section we shall construct the dual functional of $\mathscr{F}$ i.e. the functional $\mathscr{G}$: $V \to R$ satisfying

$$\sup_{v \in V} \mathscr{G}(v) = \inf_{v \in V} \mathscr{F}(v).$$

Values of $\mathscr{G}$ (or more exactly their lower approximations) will be used in the last section of the paper as d_n's.

Functional $\mathscr{F}$ can be written in the form

$$\mathscr{F}(v) = F(v) + G(\Lambda v)$$

with

$$F\colon V \to R,\ Fv = \mathscr{F}_0(\nabla v) - \int_\Omega fv \,\mathrm{d}x,\ \Lambda \in L(V, H),\ \Lambda v = v,$$

$$G\colon H \to \bar{R},\ G(p) = \int_\Omega \int_0^{p(x)} g(x, t) \,\mathrm{d}t \,\mathrm{d}x.$$

From [1-Chapter III] follows that the functional $\mathscr{L}\colon H \to \bar{R}$

$$\mathscr{L}(p) = -F^*(\Lambda' p) - G^*(-p)$$

satisfies

$$\sup_{p \in H} \mathscr{L}(p) \le \inf_{v \in V} \mathscr{F}(v). \tag{5}$$

Later (Lemma 3) we shall see that

$$\sup_{p \in H} \mathscr{L}(p) = \inf_{v \in V} \mathscr{F}(v) \tag{6}$$

holds. Since we are interested only in the value $\sup_{p \in H} \mathscr{L}(p)$, we can slightly modify the definition of $\mathscr{L}$:

$$\mathscr{L}(p) = -F^*(-\Lambda' p) - G^*(p).$$

Let us compute F^*, G^*.

Lemma 1. $F^*(v') = -\inf_{v \in V} \left\{ \int_0^1 [A_0 t\nabla v, \nabla v] \,\mathrm{d}t - \int_\Omega fv \,\mathrm{d}x - \langle v', v \rangle_V \right\},\ v' \in V'.$

Proof follows immediately from the definition of F and from the definition of the conjugate functional. ■

Lemma 2. Let us denote $g_x: R \to R$

$$g_x(t) = g(x, t).$$

Under conditions on g formulated in Section 2 it holds

$$G^*(p) = \int_\Omega \int_0^{p(x)} g_x^{-1}(s)\, \mathrm{d}s\, \mathrm{d}x, \qquad p \in H.$$

Proof. Put

$$h(x, s) = \int_0^s g(x, t)\, \mathrm{d}t.$$

Then G has the form $\mathrm{G}(p) = \int_\Omega h(x, p(x))\, \mathrm{d}x$. The assumptions of [1-Proposition IX.2.1] are satisfied and thus

$$G^*(p) = \int_\Omega \Gamma(x)\, \mathrm{d}x,$$

where $\Gamma(x) = \sup_{t \in R} \{p(x)\, t - h(x, t)\}$. If we denote $h_x: R \to R$

$$h_x(t) = h(x, t),$$

then

$$\Gamma(x) = h_x^*(p(x)).$$

The function g_x is continuous, increasing and surjective for almost all $x \in \Omega$ and thus from [3-Theorem III.4.8] it follows for almost all $x \in \Omega$ that

$$h_x^*(t) = \int_0^t g_x^{-1}(s)\, \mathrm{d}s\,.$$

Hence

$$G^*(p) = \int_\Omega \int_0^{p(x)} g_x^{-1}(s)\, \mathrm{d}s. \qquad \blacksquare$$

Using Lemma 1, Lemma 2 and the definition of $\mathscr{L}$ we have

$$\mathscr{L}(p) = \inf_{v \in V} \int_0^1 [A_0 t \nabla v, \nabla v]\, \mathrm{d}t - \int_\Omega (f - p)\, v\, \mathrm{d}x - \int_\Omega \int_0^{p(x)} g_x^{-1}(s)\, \mathrm{d}s\, \mathrm{d}x. \quad (7)$$

The following lemma describes an important property of $\mathscr{L}$.

Lemma 3. Let u be the solution of (2). Then

$$\sup_{p \in H} \mathscr{L}(p) = \mathscr{L}(g(\cdot, u)) = \mathscr{F}(u) = \inf_{v \in V} \mathscr{F}(v).$$

Before the proof of Lemma 3 we prove Lemma 4.
Lemma 4. For $v \in V$ it holds

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

Proof. The properties of g_x and [3-Theorem III.4.8] imply for almost all $x \in \Omega$

$$\int_0^{v(x)} g_x(s)\, \mathrm{d}s + \int_0^{g_x(v(x))} g_x^{-1}(s)\, \mathrm{d}s = v(x)\, g_x(v(x)).$$

Integrating this equality over Ω we obtain the assertion of Lemma 4. ∎

Proof of Lemma 3. From [3-Lemma III.4.4, Lemma III.4.11.] follows, that $\inf_{v \in V} \int_0^1 [A_0 t \nabla v, \nabla v]\, \mathrm{d}t - \int_\Omega pv\, \mathrm{d}x$ is attained at point $(\nabla' A_0 \nabla)^{-1} \Lambda' p$. Hence using (2) and the uniqueness of u we have

$$\inf_{v \in V} \int_0^1 [A_0 t \nabla v, \nabla v]\, \mathrm{d}t - \int_\Omega (f - g(\cdot, u))\, v\, \mathrm{d}x = \int_0^1 [A_0 t \nabla u, \nabla u]\, \mathrm{d}t -$$

$$- \int_\Omega (f - g(\cdot, u))\, u\, \mathrm{d}x.$$

Using (7) and Lemma 4 we obtain

$$\mathscr{L}(g(\cdot, u)) = \inf_{v \in V} \int_0^1 [A_0 t \nabla v, \nabla v]\, \mathrm{d}t - \int_\Omega (f - g(\cdot, u))\, v\, \mathrm{d}x - G^*(g(\cdot, u)) =$$

$$= \int_0^1 [A_0 t \nabla u, \nabla u]\, \mathrm{d}t - \int_\Omega (f - g(\cdot, u))\, u\, \mathrm{d}x - G^*(g(\cdot, u)) =$$

$$= \int_0^1 [A_0 t \nabla u, \nabla u]\, \mathrm{d}t - \int_\Omega (f - g(\cdot, u))\, u\, \mathrm{d}x + G(u) - \int_\Omega g(\cdot, u)\, u\, \mathrm{d}x =$$

$$= \int_0^1 [A_0 t \nabla u, \nabla u]\, \mathrm{d}t + G(u) - \int_\Omega fu\, \mathrm{d}x = \mathscr{F} u.$$

Assertion of Lemma 4 follows now from (5). ∎

We are interested only in the number $\sup_{p\in H} \mathscr{L}(p)$. From Lemma 3 it follows that it is sufficient to maximize $\mathscr{L}$ over the set $\{q \in H \mid q = g(\cdot, v)$ for some $v \in V\}$.

Thus we have the dual problem

$$\sup_{v\in V} \mathscr{G}(v), \tag{8}$$

where $\mathscr{G}\colon V \to R$

$$\mathscr{G}(v) = \mathscr{L}(g(\cdot, v)).$$

In fact the problem (8) is (considering (7)) a saddle point problem. This is unsuitable for our purposes (because we try to approximate the value $\sup_{v\in V} \mathscr{G}(v)$ from bellow). The following lemma enables us to reformulate this unsuitable saddle point "sup inf" formulation in a "sup sup" formulation.

Lemma 5. Let $p \in H$. Let us denote

$$\Phi\colon V \to R,\ \Phi(v) = \int_0^1 [A_0 t \nabla v, \nabla v]\, \mathrm{d}t - \int_\Omega (f - p)\, v\, \mathrm{d}x,$$

$$\Psi\colon U \to R,\ \Psi(w) = -\int_0^1 [Lw + R(f-p), A_0^{-1} t(Lw + R(f-p))]\, \mathrm{d}t + \\ + \int_0^1 [A_0 t A_0^{-1} 0, A_0^{-1} 0]\, \mathrm{d}t.$$

There exist unique $v_p \in V$, $w_p \in U$ such that

$$\inf_{v\in V} \Phi(v) = \Phi(v_p) = \Psi(w_p) = \sup_{w\in U} \Psi(w).$$

Moreover v_p and w_p satisfy the relation

$$Lw_p = A_0 \nabla v_p - R(f-p)$$

and they are unique solutions to problems

$$v_p \in V, \qquad \nabla' M_p \nabla v_p = 0,$$

$$w_p \in U, \qquad L' M_p^{-1} L w_p = 0,$$

where $M_p\colon \mathbf{H} \to \mathbf{H}$, $M_p(\mathbf{z}) = A_0 \mathbf{z} - R(f-p)$.

Proof. A_0 is strongly monotone, potential and thus the existence and uniqueness of v_p follows from [3-Corollary III.4.2, Theorem III.4.1.]. This v_p satisfies $\nabla' A_0 \nabla v_p = \Lambda'(f-p)$ and thus (see the property of R formulated in Introduction)

$$\nabla' M_p \nabla v_p = \nabla'(A_0 \nabla v_p - R(f-p)) = \nabla' A_0 \nabla v_p - \Lambda'(f-p) = 0. \tag{9}$$

From [3-Theorem III.4.10] we have the existence of $\mathbf{y}_p \in \operatorname{Ker} \nabla'$ such that

$$\Psi_1(\mathbf{y}_p) = \sup_{\mathbf{y} \in \operatorname{Ker} \nabla'} \Psi_1(\mathbf{y}) = \inf_{v \in V} \Phi(v) \tag{10}$$

where $\Psi_1 \colon \mathbf{H} \to R$,

$$\Psi_1(\mathbf{y}) = -\int_0^1 [\mathbf{y} + R(f-p), A_0^{-1} t(\mathbf{y} + R(f-p))] \, \mathrm{d}t + \int_0^1 [A_0 t A_0^{-1} 0, A_0^{-1} 0] \, \mathrm{d}t.$$

This $\mathbf{y}_p$ satisfies [3-Theorem III.4.10.]

$$\mathbf{y}_p = A_0 \nabla v_p - R(f-p). \tag{11}$$

Hence

$$A_0^{-1}(\mathbf{y}_p + R(f-p)) = \nabla v_p.$$

From [2], [5] if follows

$$\operatorname{Im} \nabla = \operatorname{Ker} L', \qquad \operatorname{Im} L = \operatorname{Ker} \nabla', \tag{12}$$

and thus

$$L' A_0^{-1}(\mathbf{y}_p + R(f-p)) = 0. \tag{13}$$

From (11) and from the property of R formulated in Introduction we obtain

$$\nabla' \mathbf{y}_p = \nabla' A_0 \nabla v_p - \nabla' R(f-p) = 0. \tag{14}$$

Thus from (12), (14) we have that $\mathbf{y}_p = L w_p$ for some $w_p \in U$. From (10), (12) it follows now

$$\inf_{v \in V} \Phi(v) = \Psi_1(\mathbf{y}_p) = \Psi(w_p) = \Psi_1(L w_p) =$$

$$\sup_{\mathbf{y} = \operatorname{Ker} \nabla'} \Psi_1(\mathbf{y}) = \sup_{\mathbf{y} \in \operatorname{Im} L} \Psi_1(\mathbf{y}) = \sup_{w \in U} \Psi_1(Lw) = \sup_{w \in U} \Psi(w).$$

Moreover from (13) it follows

$$L' A_0^{-1}(L w_p + R(f-p)) = L' M_p^{-1} L w_p = 0.$$

The operator M_p^{-1} is strongly monotone (and thus coercitive), continuous (see Lemma 6) and thus [3-Theorem III.2.2.] the solution $w_p \in U$ of the equation $L' M_p^{-1} L w_p = 0$ is unique. ■

Lemma 6. The operator M_p^{-1} is strongly monotone and lipschitzian:

$$[M_p^{-1}\mathbf{q}_1 - M_p^{-1}\mathbf{q}_2] \le \frac{1}{m}[\mathbf{q}_1 - \mathbf{q}_2],$$

$$[M_p\mathbf{q}_1 - M_p\mathbf{q}_2, \mathbf{q}_1 - \mathbf{q}_2] \ge \frac{m}{M^2}[\mathbf{q}_1 - \mathbf{q}_2]^2.$$

The Lipschitz and strong monotonicity constants are independent on p.

Proof. It suffices to prove that A_0^{-1} is strongly monotone and lipschitzian. Clearly $[\mathbf{q}_1 - \mathbf{q}_2] \le M[A_0^{-1}\mathbf{q}_1 - A_0^{-1}\mathbf{q}_2]$ and thus

$$[\mathbf{q}_1 - \mathbf{q}_2][A_0^{-1}\mathbf{q}_1 - A_0^{-1}\mathbf{q}_2] \ge [\mathbf{q}_1 - \mathbf{q}_2, A_0^{-1}\mathbf{q}_1 - A_0^{-1}\mathbf{q}_2] \ge m[A_0^{-1}\mathbf{q}_1 - A_0^{-1}\mathbf{q}_2]^2 \ge$$

$$\ge \frac{m}{M^2}[\mathbf{q}_1 - \mathbf{q}_2]^2. \quad \blacksquare$$

Using Lemma 3, Lemma 5 and the definition of $\mathscr{G}$ we can summarize the results of this section.

Proposition 2. The functional $\mathscr{G}$: $V \to R$ defined by

$$\mathscr{G}(v) = \sup_{w \in U} - \int_0^1 [Lw + R(f - g(\cdot, v)), A_0^{-1}t(Lw + R(f - g(\cdot, v)))]\,\mathrm{d}t + \\ + \int_0^1 [A_0 t A_0^{-1}0, A_0^{-1}0]\,\mathrm{d}t - \int_\Omega \int_0^{g_x(v(x))} g_x^{-1}(s)\,\mathrm{d}s\,\mathrm{d}x \tag{15}$$

satisfies

$$\mathscr{G}(v) = (\mathscr{L}(g., v)) = \inf_{w \in V} \int_0^1 [A_0 t\nabla w, \nabla w] - \int_\Omega (f - g(\cdot, v)\, w\,\mathrm{d}x - \\ - \int_\Omega \int_0^{g(x, v(x))} g_x^{-1}(s)\,\mathrm{d}s\,\mathrm{d}x, \; v \in V,$$

$$\sup_{v \in V} \mathscr{G}(v) = \inf_{v \in V} \mathscr{F}(v).$$

The following section deals with the problem of approximation of $\sup_{v \in V} \mathscr{G}(v)$ from bellow using the last formulation of $\mathscr{G}$.

4. Approximate solution of the dual problem, a posteriori error estimates

It is clear (Lemma 3) that supremum in (8) is attained at point u. However, this point is not known explicitly. Our approach is to approximate the value (8) using the minimizing sequence u_n, $n = 1, 2, \ldots$ of $\mathscr{F}$.

Theorem 1. Let u, $\mathscr{F}$, $\mathscr{G}$ be defined by (2), (3), (15) and $u_n \in V$, $n = 1, 2, \dots$ be a minimizing sequence of $\mathscr{F}$. Let $s_n = \mathscr{G}(u_n)$, $n = 1, 2, \dots$. Then

$$\| u_n - u \|^2 \leq \frac{2}{m} (\mathscr{F}(u_n) - s_n) \to 0.$$

Proof. Let us denote $A\colon V \to V'$, $A = \nabla' A_0 \nabla$. Analogously to the proof of Lemma 3

$$s_n = \inf_{v \in V} \int_0^1 [A_0 t \nabla v, \nabla v] \,\mathrm{d}t - \int_\Omega (f - g(\cdot, u_n))\, v \,\mathrm{d}x - G^*(g(\cdot, u_n)) =$$

$$= \int_0^1 [A_0 t \nabla A^{-1} \Lambda'(f - g(\cdot, u_n)), \nabla A^{-1} \Lambda'(f - g(\cdot, u_n))] \,\mathrm{d}t -$$

$$- \int_\Omega (f - g(\cdot, u_n))\, A^{-1} \Lambda'(f - g(\cdot, u_n)) \,\mathrm{d}x - G^*(g(\cdot, u_n)).$$

The operator A^{-1} is continuous [3-Theorem III.2.2.]. From Lemma 4 it follows that $G^*(g(\cdot, u_n)) \to G^*(g(\cdot, u))$ because $G(\cdot, v)$ and $\int_\Omega v g(\cdot, v) \,\mathrm{d}x$ are continuous in v over V. Thus using (2) and Lemma 4 we obtain

$$s_n \to \int_0^1 [A_0 t \nabla u, \nabla u] \,\mathrm{d}t - \int_\Omega (f - g(\cdot, u))\, u \,\mathrm{d}x + G(u) - \int_\Omega u g(\cdot, u) \,\mathrm{d}x = \mathscr{F}(u).$$

The rest of the proof follows from the inequality $s_n \leq \sup_{v \in V} \mathscr{G}(v) = \mathscr{F}(u)$ and from Corollary 1. ■

Unfortunately the values of $\mathscr{G}$ cannot be computed explicitly. However the "sup" formulation of $\mathscr{G}$ (see (15)) enables us to approximate them from bellow.

Theorem 2. Let u, $\mathscr{F}$ be defined by (2), (3) and u_n, $n = 1, 2, \dots$ be a minimizing sequence of $\mathscr{F}$. Let w_n, $n = 1, 2, \dots$ be the unique solution to the problem

$$w_n \in U_n, \qquad L_n' B_n L_n w_n = 0, \tag{16}$$

where $B_n = M^{-1}_{g(\cdot, u_n)}$. (For definition of $M_{g(\cdot, u_n)}$ see Lemma 5.)

Then for r_n, $n = 1, 2, \dots$ defined by

$$r_n = - \int_0^1 [L w_n + R(f - g(\cdot, u_n)), A_0^{-1} t (L w_n + R(f - g(\cdot, u_n)))] \,\mathrm{d}t +$$

$$+ \int_0^1 [A_0 t A_0^{-1} 0, A_0^{-1} 0] \,\mathrm{d}t \qquad - \int_\Omega \int_0^{g(x, u_n(x))} g_x^{-1}(t) \,\mathrm{d}t \,\mathrm{d}x$$

it holds

$$\| u_n - u \|^2 \le \frac{0}{m} (\mathscr{F}(u_n) - r_n) \to 0.$$

Proof. Let the value s_n (see Theorem 1) be attained at $v_n \in U$ i.e.

$$\mathscr{G}(u_n) = - \int_0^1 [Lv_n + R(f - g(\cdot, u_n)), A_0^{-1} t(Lv_n + R(f - g(\cdot, u_n)))] \,dt +$$

$$+ \int_0^1 [A_0 t A_0^{-1} 0, A_0^{-1} 0] \,dt - \int_\Omega \int_0^{g(x, u_n(x))} g_x^{-1}(t) \,dt \,dx$$

We will show that $v_n - w_n \to 0$ in U. First we prove

Lemma 7. $v_n \to v_0$ in U, where v_0 is the (unique) solution of the problem

$$v_0 \in U, \qquad L' A_0^{-1}(Lv_0 + R(f - g(\cdot, u))) = 0.$$

(u is the solution of (2)).

Proof. We know from Lemma 5 that $v_n \in U$ is the unique solution to the problem

$$v_n \in U,\ L' B_n L v_n = 0. \tag{17}$$

Let the operator S: $U \times V \to U$ be defined by

$$S(v, y) = v - tJ^{-1}(L' A_0^{-1}(Lv + R(f - g(\cdot, y)))), \qquad (0 < t < 2m^3/M^2).$$

We shall show that S is uniformly (relative to y) contractive in v and continuous in y and thus the uniform contractivity theorem (e.g. [6] Theorem XVI.1.3.) can be used.

From Lemma 6 it follows, that for $y \in V$, the operator C_y: $U \to U'$,

$$C_y v = L' M_{g(\cdot, y)}^{-1} L v$$

is strongly monotone and lipschitzian with the Lipschitz and strong monotonicity constants independent on y:

$$\forall v_1, v_2 \in U \colon \langle C_y v_1 - C_y v_2, v_1 - v_2 \rangle_U = [M_{g(\cdot, y)}^{-1} L v_1 - M_{g(\cdot, y)}^{-1} L v_2, Lv_1 - Lv_2] \ge$$

$$\ge \frac{m}{M^2} \| v_1 - v_2 \|^2,$$

$$\forall v_1, v_2 \in U \colon \| C_y v_1 - C_y v_2 \|_{U'} = \| L'(M_{g(\cdot, y)}^{-1} L v_1 - M_{g(\cdot, y)}^{-1} L v_2) \|_{U'} \le$$

$$\le \| L' \|_{L(\mathrm{H}, U')} \frac{1}{m} [L(v_1 - v_2)] = \frac{1}{m} \| v_1 - v_2 \|,$$

since $\| L \|_{L(U, \mathrm{H})} = \| L' \|_{L(\mathrm{H}, U')} = 1$, $[Lv] = \| v \|$, $\forall v \in U$.

The properties of J [3-Lemma I.6.2] imply now

$$\forall v_1, v_2 \in U: ((J^{-1}C_y v_1 - J^{-1}C_y v_2, v_1 - v_2)) = \langle C_y v_1 - C_y v_2, v_1 - v_2\rangle_U \geq$$

$$\geq \frac{m}{M^2} \| v_1 - v_2 \|^2,$$

$$\forall v_1, v_2 \in U: \| J^{-1}C_y v_1 - J^{-1}C_y v_2 \| = \| C_y v_1 - C_y v_2 \|_{U'} \leq \frac{1}{m} \| v_1 - v_2 \| .$$

Thus the operator $J^{-1}C_y$ satisfies the assumptions of [3-Lemma III.3.1.]. Therefore, for $y \in V$, the operator $S_y^{(1)}: U \to U$,

$$S_y^{(1)}(v) = S(v, y) \qquad (= v - tJ^{-1}C_y v)$$

is contractive in v for $0 < t < 2m^3/M^2$. The contractivity constant is independent on y. The operator $S_v^{(2)}(y): V \to U$

$$S_v^{(2)}(y) = S(v, y)$$

is continuous in y for $v \in U$ because g satisfies the growth conditions (see Section 2) and R, A_0^{-1}, L' are continuous operators.

It is clear (Lemma 5), that the fixed point of $S_{u_n}^{(1)}$ is v_n. The uniform contractivity theorem implies

$$u_n \to u \Rightarrow v_n \to v_0,$$

where v_0 is the fixed point of $S_u^{(1)}$. Hence $L'A_0^{-1}(Lv_0 + R(f - g(\cdot, u))) = 0$. ∎

From (16), (17) it follows that w_n is the Galerkin approximation of v_n. The strong monotonicity and Lipschitz constants of B_n, $n = 1, 2, \ldots$ are independent on n and thus [3-Theorem III.3.3.]

$$\| v_n - w_n \| \leq C \| P_n v_n - v_n \|, \; n = 1, 2, \ldots .$$

The constant C is independent on n. Hence using the property of P_n, $n = 1, 2, \ldots$ formulated in Introduction and the well known property $\| P_n \|_{L(U, U_n)} = 1$, $n = 1, 2, \ldots$ we obtain

$$\| w_n - v_n \| \leq C \| P_n v_n - v_n \| = C \| P_n v_n - P_n v_0 + P_n v_0 - v_0 + v_0 - v_n \| \leq$$

$$\leq C \| P_n \|_{L(U, U_n)} \| v_n - v_0 \| + C \| P_n v_0 - v_0 \| + C \| v_0 - v_n \| \to 0.$$

The continuity of Ψ (for the definition of Ψ see Lemma 5) implies now $\lim_{n \to x} r_n = \lim_{n \to \infty} s_n = \mathscr{F}(u)$. The rest of the proof follows from Corollary 1. ∎

The computation of v_n, $n = 1, 2, \ldots$ in Theorem 2 requires the solution of a nonlinear equation in a finite-dimensional space. The following theorem requires only a solution of a linear equation in finite-dimensional space.

Theorem 3. Let u, $\mathscr{F}$ be defined by (2), (3) and $u_n \in V$, $n = 1, 2, \ldots$ be a minimizing sequence of $\mathscr{F}$. Let $z_1 \in U_1$, the sequence z_n, $n = 2, 3, \ldots$ be defined by

$$z_n = D_n z_{n-1},$$

where D_n, $n = 2, 3, \ldots$ is the operator $D_n \colon U_n \to U_n$

$$D_n w = w - tJ_n^{-1}(L_n' B_n L_n)\, w, \qquad (0 < t < 2m^3/M^2). \tag{18}$$

(For definition of B_n [see Theorem 2].) Then for the sequence d_n, $n = 1, 2, \ldots$ defined by

$$d_n = -\int_0^1 [Lz_n + R(f - g(\cdot, u_n)), A_0^{-1} t(Lz_n + R(f - g(\cdot, u_n)))]\, \mathrm{d}t +$$

$$+ \int_0^1 [A_0^{-1} t A_0^{-1} 0, A_0^{-1} 0]\, \mathrm{d}t - \int_\Omega \int_0^{g_x(u_n(x))} g_x^{-1}(t)\, \mathrm{d}t\, \mathrm{d}x$$

it holds

$$\| u_n - u \|_2 \leq \frac{2}{m} (\mathscr{F}(u_n) - d_n) \to 0.$$

Proof. Repeating the argument of the proof of contractivity of $S_y^{(1)}$ in Theorem 2 (replacing U, J, L by U_n, $J_n L_n$) we can show the contractivity of D_n, $n = 1, 2, \ldots$. The contractivity constant is again independent on n.

The fixed point of D_n is w_n, $n = 1, 2, \ldots$. (For the definition of w_n see Theorem 2.) Thus according to [3-Lemma III.3.2.] the sequence z_n, $n = 1, 2, \ldots$ defined by (18) tends (in U) to $v_0 (= \lim_{n \to \infty} w_n)$. The rest of the proof follows from the continuity of Ψ and from Corollary 1. ■

Remark 3. The computation of z_n, $n = 2, 3, \ldots$ requires in each step solution of a linear equation in a finite-dimensional space, namely the solution of the equation

$$J_n z_n = J_n z_{n-1} - tL_n' B_n L_n z_{n-1}, \qquad z_n \in U_n. \tag{19}$$

Let $a_1, \ldots, a_n$ be the basis of U_n. Then (19) is equivalent to

$$\langle J_n z_n, a_j \rangle_U \equiv [Lz_n, La_j] = [Lz_{n-1}, La_j] - t[B_n L_n z_{n-1}, L_n a_j], \qquad j = 1, 2, \ldots n.$$

Thus

$$z_n = \sum_{i=1}^{n} \alpha_i a_i,$$

where $(\alpha_1, \ldots, \alpha_n) \in R^n$ is the unique solution of the system of linear algebraic equations

$$\sum_{i=1}^{n} \alpha_i((a_i, a_j)) = ((z_{n-1}, a_j)) - t[B_n L z_{n-1}, L a_j], \qquad j = 1, 2, \ldots n. \tag{20}$$

(We used the identity $[Lv, Lw] = ((v, w)))$ The matrix of the system in (20) is the Gramm matrix and thus it is positive definite.

REFERENCES

1. Ekeland, I., Temam, R.: Convex Analysis and Variational problems, North-Holland, Amsterdam 1976.
2. Gajewski, H., Grögär, K.: Konjugierte Operatoren und a-posteriori Fehlerabschatzungen. Math. Nachr. 73, 315—333 (1976).
3. Gajewski, H., Gröger, K.: Zacharias, K., Nichtlineare Operatorgleichungen und Operatordifferentialgleichungen, Akademie-Verlag Berlin 1974 (Russian Mir Moskva 1978).
4. Haslinger, J., Hlaváček, I.: Convergence of a finite element method based on the dual variational formulation, Aplikace matematiky 21, 43—65 (1976).
5. Křížek, M.: Conforming equilibrium finite element methods for some elliptic plane problems, RAIRO Anal. Numer. 17, 35—65 (1983).
6. Kantorowitch, L. V., Akilow, G. P.: Functional Analysis, Nauka, Moscow 1984 (Russian).
7. Shampine, L. F.: Error Bounds and Variational Methods for Nonlinear Boundary Value problems, Numer. Math. 12, 410—415 (1968).
8. Vacek, J.: Dual variational principles for an elliptic partial differential equation, Aplikace Matematiky 21, 5—27 (1976).
9. Weisz, J.: A posteriori error estimate of approximate solutions to a mildly nonlinear elliptic boundary value problem, CMUC 31, 315—322 (1990).

Author's adress:

Received 16. 11. 1989

Juraj Weisz
Ústav aplikovanej matematiky MFF UK
Mlynská dolina,
842 15 Bratislava,

РЕЗЮМЕ

АПОСТЕРИОРНАЯ ОЦЕНКА ПОГРЕШНОСТИ ПРИБЛИЗИТЕЛЬНЫХ РЕШЕНИЙ НЕЛИНЕЙНОЙ ЭЛЛИПТИЧЕСКОЙ КРАЕВОЙ ЗАДАЧИ

Юрай Веисз, Братислава

Статья посвящена конструкции апостериорных оценок погрешности приблизительных решений нелинейной эллиптической краевой задачи с краевыми условиями Дирихлета

$$-\sum_{i=1}^{2} \frac{\partial}{\partial x_i} (a_i(x, \nabla u)) + g(x, u) = f(x), \qquad \text{in } \Omega$$

$$u|_{\partial\Omega} = 0.$$

Показана сходимость конструированной оценки к действительной ошибке. Для вычисления оценки погрешности надо решить систему линейных алгебраических уравнений.

SÚHRN

APOSTERIÓRNY ODHAD CHYBY PRIBLIŽNÝCH RIEŠENÍ NELINEÁRNEJ ELIPTICKEJ OKRAJOVEJ ÚLOHY

Juraj Weisz, Bratislava

Práca sa zaoberá konštrukciou aposteriórnych odhadov presnosti približných riešení nelineárnej eliptickej okrajovej úlohy s Dirichletovou okrajovou podmienkou

$$-\sum_{i=1}^{2} \frac{\partial}{\partial x_i} a_i(x, \nabla u) + g(x, u) = f(x),$$

$$u|_{\partial\Omega} = 0.$$

Dokazuje sa konvergencia odhadu skonštruovaného v práci ku skutočnej chybe približného riešenia. Výpočet odhadu chyby vyžaduje riešenie systému lineárnych algebraických rovníc.